THE LOUSE FLY-ARSENOPHONUS ARTHROPODICUS ASSOCIATION: DEVELOPMENT OF A NEW MODEL SYSTEM FOR THE STUDY OF INSECT-BACTERIAL ENDOSYMBIOSES

by

Kari Lyn Smith

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biology

The University of Utah

August 2012

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Kari Lyn Smith**

has been approved by the following supervisory committee members:

Colin Dale	, Chair	**June 18, 2012** Date Approved
Dale Clayton	, Member	**June 18, 2012** Date Approved
Maria-Denise Dearing	, Member	**June 18, 2012** Date Approved
Jon Seger	, Member	**June 18, 2012** Date Approved
Robert Weiss	, Member	**June 18, 2012** Date Approved

and by **Neil Vickers**, Chair of the Department of **Biology**

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

There are many bacteria that associate with insects in a mutualistic manner and offer their hosts distinct fitness advantages, and thus have likely played an important role in shaping the ecology and evolution of insects. Therefore, there is much interest in understanding how these relationships are initiated and maintained and the molecular mechanisms involved in this process, as well as interest in developing symbionts as platforms for paratransgenesis to combat disease transmission by insect hosts. However, this research has been hampered by having only a limited number of systems to work with, due to the difficulties in isolating and modifying bacterial symbionts in the lab. In this dissertation, I present my work in developing a recently described insect-bacterial symbiosis, that of the louse fly, *Pseudolynchia canariensis*, and its bacterial symbiont, *Candidatus* Arsenophonus arthropodicus, into a new model system with which to investigate the mechanisms and evolution of symbiosis. This included generating and analyzing the complete genome sequence of *Ca.* A. arthropodicus, which provided some evidence that *Ca.* A. arthropodicus has become recently associated with insects and may have evolved from an ancestor that was an insect pathogen. Additionally, I describe the development of methods for genetic modification of this bacterial symbiont and for introducing recombinant symbionts into louse fly hosts, as well as a new microinjection technique that enables the complete replacement of native symbionts with recombinant symbionts. With the generation of the symbiont genome sequence along with strategies

for engineering recombinant symbionts and establishing them in an insect host, this work provides an interesting new system with which to investigate the function of specific genes in symbiosis as well as a promising new avenue of research involving paratransgenesis.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER 1

INTRODUCTION

Bacteria have long been known to interact with a wide range of eukaryotic hosts, though historically, much study in this area has focused on organisms that cause disease. However, there are many bacteria that associate with eukaryotic organisms in a mutualistic manner and offer their hosts distinct fitness advantages. A wide variety of insects are known to harbor mutualistic bacterial symbionts (Buchner, 1965), which maximize host fitness in exchange for a constant, protected environment. These insect-bacterial symbioses may have a very ancient origin of association, consisting of "primary" symbionts that have become highly specialized to the insect host environment and are often housed in specific host-derived cells termed bacteriocytes that make up an organ called the bacteriome (Dale and Moran, 2006; Douglas, 2011). Primary symbionts are often obligately required by the insect host and serve to supplement the host diet with various nutrients, allowing it to subsist on a food source that may not be nutritionally complete (Dale and Moran, 2006). For example, the ancient symbiont *Buchnera aphidicola* resides within aphid hosts and serves to supplement the aphid's diet of plant sap with vital amino acids that are lacking (Douglas, 1998). *Wigglesworthia glossinidia* is another ancient primary symbiont associated with tsetse flies that plays a role in supplementing the tsetse's vertebrate blood diet with B vitamins and aids in host digestion of the blood meal (Akman et al., 2002; Pais et al., 2008). Without their primary

symbionts, insect hosts generally suffer fitness effects such as reduced fecundity or sterility and decreased life span (Nogge 1976, 1981; Douglas, 1998; Heddi et al., 1999; Pais et al., 2008).

Mutualistic insect symbionts are typically vertically transmitted directly from mother to offspring, often transovarially or through ingestion of symbionts by offspring during a specific life stage (Bright and Bulgheresi, 2010). For example, *B. aphidicola* is transmitted via direct entry by symbionts into developing eggs or embryos (Wilkinson et al., 2003), while *W. glossinidia* is introduced into tsetse fly larvae through *in utero* feeding of the larvae on maternal milk gland secretions (Attardo et al., 2008). Due to this strict vertical transmission and their long evolutionary history of association with their insect hosts, primary symbionts often show a pattern of cospeciation with their hosts, evidenced by congruent branching patterns of symbiont and host phylogenies (Funk et al., 2000; Hosokawa et al., 2006).

In addition to ancient symbionts, insects may also harbor one to a few bacterial symbionts with which they share a recent origin of association, that are often referred to as "secondary" or "facultative" symbionts. These symbionts may reside in bacteriocytes, although they often are not sequestered solely to these cells and may be found in a variety of host tissues including the fat body, hemolymph and reproductive tissues (Cheng and Aksoy, 1999; Dale et al., 2006; Oliver et al., 2010). Recently derived symbionts may provide a variety of benefits to their hosts, including diet supplementation or they may function in some capacity unrelated to host nutrition. For example, the secondary symbiont of tsetse flies, *Sodalis glossinidius*, retains genes for the production of B vitamins and may play a role in supplementing its host's blood diet (Nogge, 1981; Toh et

al., 2006), and recent symbionts of aphids may be involved in host plant specialization (Tsuchida et al., 2004; Oliver et al., 2010). However, secondary symbionts have been implicated in a range of other functions as well, such as increasing host resistance to parasites or environmental stresses. For example, in aphids, the symbiont *Regiella insecticola* helps to protect its host from fungal pathogen infection (Scarborough et al., 2005), while the symbiont *Hamiltonella defensa* increases aphid resistance to parasitoid wasps (Oliver et al., 2009, 2010), and the presence of *Serratia symbiotica* enhances aphid host survival and fecundity under heat stress (Montllor et al., 2002; Russell and Moran et al., 2006). However, in many other insect associations, the roles of recently derived symbionts remain unclear.

Recent mutualistic symbionts are often vertically transmitted in the same manner as primary symbionts, although they may undergo horizontal transfer between different insect hosts on occasion. Unlike ancient symbionts, the phylogenies of recently derived symbionts are not concordant with their host phylogenies (Russell et al., 2003; Dale and Moran, 2006). Additionally, some closely related symbionts have been identified in a wide range of insect hosts, indicating that these symbionts have been acquired independently from the environment or through horizontal transfer (Novakova and Hypsa, 2007; Moran et al., 2008; Novakova et al., 2009).

With the advent of new sequencing technologies, there has been a rapid increase in the number of bacterial genome sequencing projects that have been completed, including many insect symbionts, with both ancient and recent origins of association with insects. This genomic information has provided much insight into the evolution and functional basis of insect-bacterial symbioses. What has become clear is that often the

bacterial symbionts of insects embark on a trajectory of genome degeneration and size reduction once they become associated with an insect host (Moran et al., 2008). This is predicted to be based on the fact that, once host-associated, the bacteria inhabit a very well-protected niche that is more static than the environments encountered by their free-living relatives. Therefore, there is relaxed selection on genes that may not be necessary in the new insect niche and these genes may quickly be inactivated in symbionts via large deletions, frameshifting mutations or altered start/stop codons (Burke and Moran, 2011). Additionally, genes under relaxed selection may serve as an insertion point for transposons or phage elements and there is often an expansion of these elements in recently derived symbionts (Belda et al., 2010; McCutcheon and Moran, 2012). Over time, repetitive DNA elements and inactivated genes are deleted from the genome to the point that only genes required for the symbiotic association are maintained and the genome size is greatly reduced (McCutcheon and Moran, 2012). For example, the smallest known bacterial genomes belong to primary symbionts of insects, including the psyllid symbiont *Candidatus* Carsonella rudii (160 kb; Nakabachi et al., 2006) and the mealybug symbiont, *Candidatus* Tremblaya princeps (139 kb; Lopez-Madrigal et al., 2011). These symbionts both retain genes required for the biosynthesis of amino acids that are anticipated to be lacking in the insect host diet even though they have lost many other pathways predicted to be essential for bacteria and must rely on their hosts or other symbionts for a variety of basic metabolic intermediates and cofactors (Nakabachi et al., 2006; Lopez-Madrigal et al., 2011; McCutcheon and Moran, 2012).

The transition from a free-living bacterium with a large genome to a host-associated obligate symbiont with a small genome is a dynamic process and there is much

interest in understanding the time course and mechanisms by which this occurs. Due to their more recent origin of association, the genomes of secondary symbionts tend to represent an intermediate stage of genome degeneration, in which many genes have acquired inactivating mutations, such as those creating frameshifts and premature stop codons, but have not yet been fully deleted from the genome (Belda et al., 2010; Burke and Moran, 2011). This provides an interesting snapshot of the reductive process and these intermediate genomes may provide information on the mechanisms involved in undergoing a change from a free-living organism to an intracellular, obligate symbiont.

The genome degeneration that occurs over time in insect symbionts has made study of the symbionts in the lab very difficult. The loss of many genes maintained by free-living bacteria to survive in a variety of environments makes symbionts very fastidious and difficult to isolate in culture, especially in the case of obligate primary symbionts. To date, only a few insect symbionts have been isolated in pure culture, and those that have proved amenable to culture are more recently derived symbionts that still maintain a larger gene set (Dale and Maudlin, 1999; Dale et al., 2006; Sabri et al., 2010). Culture and genetic manipulation of insect symbionts is a very important tool needed to test predictions about gene function and utility in symbioses and gain a better understanding of the molecular mechanisms required to initiate and maintain these associations. One insect symbiont that has been isolated in culture is that of tsetse flies, *S. glossinidius,* and recent work has involved genetic modification of this symbiont and introduction into insect hosts (Dale and Maudlin, 1999; Pontes and Dale, 2011; Pontes et al., 2011). However, the genetic techniques available for use in symbionts still lag behind those available for insect pathogens. There is a need for additional symbiont

study systems which are amenable to culture and genetic manipulation for functional studies *in vitro* and *in vivo*.

An ideal study system for insect-bacterial symbioses would consist not only of a bacterial symbiont that can be cultured and genetically manipulated, but also a host insect that can be cultured and infected with genetic variants of symbionts to observe effects. Since many insects harbor multiple mutualistic symbionts that are transmitted maternally during insect reproduction, there are often no life stages during which these insects are aposymbiotic. This imposes a challenge in the establishment of a population of genetically modified symbionts and of observing the effects of genetic variants on the association. Recent attempts to establish tsetse flies carrying recombinant *S. glossinidius* have addressed this issue by first treating insects with antibiotics to reduce or eliminate wild type *S. glossinidius* prior to the introduction of recombinant symbionts (Weiss et al., 2006). However, antibiotics tend to lack specificity and many insects, including tsetse flies, are known to harbor multiple mutualistic symbionts (Dale and Moran, 2006). This makes it difficult to selectively eliminate a particular symbiont without affecting other members of the symbiotic flora, thereby compromising the fitness of the host insect. Thus, new methodologies are needed for introducing recombinant bacterial symbionts into insect hosts in order to investigate the molecular mechanisms of symbiosis.

The work presented here focuses on the development of the association between the bacterial symbiont *Candidatus* Arsenophonus arthropodicus, and its louse fly host, *Pseudolynchia canariensis*, into a tractable system with which to study insect-bacterial symbioses. Closely related *Arsenophonus* species have been identified as symbionts in a wide range of distantly related insect hosts, such as triatomines (Hypsa and Dale, 1997),

ticks (Grindle et al., 2003), whiteflies (Thao and Baumann, 2004), hippoboscid flies (Trowbridge et al., 2006), and parasitoid wasps (Gherna et al., 1991), among a number of other insects (Novakova et al., 2009) and these *Arsenophonus* symbionts interact with their hosts in a variety of ways. These symbionts may be bacteriome-associated or distributed throughout the host body, vertically or horizontally transmitted, and they may maintain mutualistic relationships with insect hosts or behave as reproductive parasites that manipulate host reproduction, in a manner similar to the lifestyle of many *Wolbachia* species (Wilkes et al., 2011). Due to their widespread distribution, it seems unlikely that the general role of *Arsenophonus* symbionts in insect symbioses is diet supplementation, given that identified insect hosts subsist on diverse diets with differing nutrient availabilities, although it is unclear at this point what the functions of these symbionts might be in these varied associations.

Candidatus Arsenophonus arthropodicus, the pigeon louse fly symbiont, displays traits common to other mutualistic endosymbionts of insects, including vertical transmission to offspring, and louse fly hosts do not exhibit any reproductive effects due to this association (Dale et al., 2006). This symbiont has been isolated in axenic culture and has proved amenable to genetic transformation with broad host-range plasmids (Dale et al., 2006). In addition, louse fly hosts can be maintained in the lab by culturing them on their native pigeon hosts. This provides a useful prospective study system for investigating the role of this bacterium in various symbioses and the mechanisms involved in initiating and sustaining successful infections in insect hosts.

Here, the complete genome sequene for *Ca.* A. arthropodicus is presented and compared to other bacterial genomes. Understanding the gene inventory may help

elucidate the function of this symbiont in its louse fly host as well as provide an additional genome for comparison to gain a better understanding of insect symbiont evolution. The *Ca.* A. arthropodicus genome sequence is then compared to the draft sequence of another *Arsenophonus* symbiont, *Arsenophonus nasoniae* (Darby et al., 2010), a symbiont of the parasitoid wasp, *Nasonia vitripennis* (Gherna et al., 1991). *Arsenophonus nasoniae* has a different lifestyle than *Ca.* A. arthropodicus, with a male killing phenotype in its wasp host and it can undergo both vertical and horizontal transmission (Gherna et al., 1991; Duron et al., 2010). The gene inventories of both symbionts indicate a recent origin of association with insects and contain evidence of a transition from a pathogenic ancestor.

The availability of the complete gene inventory for *Ca.* A. arthropodicus provides a variety of target genes of interest to begin to investigate using molecular genetic techniques. In this work, genetic modification of *Ca.* A. arthropodicus by homologous recombination mediated by the lambda Red recombineering system (Datsenko and Wanner, 2000) is described, as well as the introduction of modified bacteria into louse fly hosts via microinjection. Recombinant bacteria were able to successfully initiate an infection in louse flies, and underwent vertical transmission to offspring when injected into hosts during the pupal stage. Additionally, aposymbiotic louse flies were obtained as a result of the microinjection procedure, providing a means to investigate possible roles of *Ca.* A. arthropodicus in this association under varied environmental conditions.

Given their broad host distribution, understanding the functions of this symbiont in the louse fly and the molecular mechanisms of infection might provide some insight into the ability of this genus to infect such a wide range of distantly related insects.

Additionally, there is much interest in developing insect platforms for studies of paratransgenesis, which entail using bacterial symbionts to express transgenes in insects to reduce their capability to transmit parasites in the wild (Coutinho-Abreu et al., 2010). Thus, the louse fly - *Ca.* A. arthropodicus association provides a promising new system with which to investigate the mechanisms of insect-bacterial symbioses as well as techniques of paratransgenesis for disease control.

References

Akman, L., A. Yamashita, H. Watanabe, K. Oshima, T. Shiba, M. Hattori, and S. Aksoy. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia. *Nat. Genet.* 32:402-407.

Attardo, G.M., C. Lohs, A. Heddi, U.H. Alam, S. Yildirim, and S. Aksoy. 2008. Analysis of milk gland structure and function in Glossina morsitans: milk protein production, symbiont populations and fecundity. *J. Insect Physiol.* 54:1236-1242.

Bright, M., and S. Bulgheresi. 2010. A complex journey: transmission of microbial symbionts. *Nat. Rev. Microbiol.* 8:218-230.

Buchner, P. 1965. Endosymbiosis of Animals with Plant Microorganisms. John Wiley, New York. 909 pp.

Burke, G.R., and N.A. Moran. 2011. Massive genomic decay in Serratia symbiotica, a recently evolved symbiont of aphids. *Genome Biol. Evol.* 3:195-208.

Cheng, Q., and S. Aksoy. 1999. Tissue tropism, transmission and expression of foreign genes in vivo in midgut symbionts of tsetse flies. *Insect Mol. Biol.* 8:125-132.

Coutinho-Abreu, I.V., K.Y. Zhu, and M. Ramalho-Ortigao. 2010. Transgenesis and paratransgenesis to control insect-borne diseases: current status and future challenges. *Parasitol. Int.* 59:1-8.

Dale, C., M. Beeton, C. Harbison, T. Jones, and M. Pontes. 2006. Isolation, pure culture, and characterization of "Candidatus Arsenophonus arthropodicus," an intracellular secondary endosymbiont from the hippoboscid louse fly Pseudolynchia canariensis. *Appl. Environ. Microbiol.* 72:2997-3004.

Dale, C., and I. Maudlin. 1999. Sodalis gen. nov. and Sodalis glossinidius sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly Glossina morsitans

morsitans. *Int. J. Syst. Bacteriol.* 49 Pt 1:267-275.

Dale, C., and N.A. Moran. 2006. Molecular interactions between bacterial symbionts and their hosts. *Cell.* 126:453-465.

Darby, A.C., J.H. Choi, T. Wilkes, M.A. Hughes, J.H. Werren, G.D. Hurst, and J.K. Colbourne. 2010. Characteristics of the genome of Arsenophonus nasoniae, son-killer bacterium of the wasp Nasonia. *Insect Mol. Biol.* 19 Suppl 1:75-89.

Datsenko, K.A., and B.L. Wanner. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. USA.* 97:6640-6645.

Douglas, A.E. 1998. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera. *Annu. Rev. Entomol.* 43:17-37.

Duron, O., T.E. Wilkes, and G.D. Hurst. 2010. Interspecific transmission of a male-killing bacterium on an ecological timescale. *Ecol. Lett.* 13:1139-1148.

Funk, D.J., L. Helbling, J.J. Wernegreen, and N.A. Moran. 2000. Intraspecific phylogenetic congruence among multiple symbiont genomes. *Proc. Biol. Sci.* 267:2517-2521.

Gherna, R.L., Werren, J.H., Weisburg, W., Cote, R., Woese, C.R., Mandelco, L., and D.J. Brenner. 1991. Arsenophonus nasoniae gen. nov., sp. nov., the causative agent of the son-killer trait in the parasitic wasp Nasonia vitripennis. *Int. J. Syst. Bacteriol.* 41:563-565.

Grindle, N., J.J. Tyner, K. Clay, and C. Fuqua. 2003. Identification of Arsenophonus-type bacteria from the dog tick Dermacentor variabilis. *J. Invertebr. Pathol.* 83:264-266.

Heddi, A., A.M. Grenier, C. Khatchadourian, H. Charles, and P. Nardon. 1999. Four intracellular genomes direct weevil biology: nuclear, mitochondrial, principal endosymbiont, and Wolbachia. *Proc. Natl. Acad. Sci. USA.* 96:6814-6819.

Hosokawa, T., Y. Kikuchi, N. Nikoh, M. Shimada, and T. Fukatsu. 2006. Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. *PLoS Biol.* 4:e337.

Hypsa, V., and C. Dale. 1997. In vitro culture and phylogenetic analysis of "*Candidatus* Arsenophonus triatominarum," and intracellular bacterium from the triatomine bug, *Triatoma infestans*. *Int. J. Syst. Bacteriol.* 47:1140-1144.

Lopez-Madrigal, S., A. Latorre, M. Porcar, A. Moya, and R. Gil. 2011. Complete genome sequence of "Candidatus Tremblaya princeps" strain PCVAL, an intriguing

translational machine below the living-cell status. *J. Bacteriol.* 193:5587-5588.

McCutcheon, J.P., and N.A. Moran. 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol.* 10:13-26.

Montllor, C.B., Maxmen, A., and A.H. Purcell. 2002. Facultative bacterial endosymbionts benefit pea aphids Acyrthosiphon pisum under heat stress. *Ecol. Entomol.* 27:189-195.

Moran, N.A., J.P. McCutcheon, and A. Nakabachi. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* 42:165-190.

Nakabachi, A., A. Yamashita, H. Toh, H. Ishikawa, H.E. Dunbar, N.A. Moran, and M. Hattori. 2006. The 160-kilobase genome of the bacterial endosymbiont Carsonella. *Science*. 314:267.

Nogge, G. 1976. Sterility in tsetse flies (Glossina morsitans Westwood) caused by loss of symbionts. *Experientia*. 32:995-996.

Nogge, G. 1981. Significance of symbionts for the maintenance of an optional nutritional state for successful reproduction in hematophagous arthropods. *Parasitol.* 82:101-104.

Novakova, E., and V. Hypsa. 2007. A new Sodalis lineage from bloodsucking fly Craterina melbae (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont Sodalis glossinidius. *FEMS Microbiol. Lett.* 269:131-135.

Novakova, E., V. Hypsa, and N.A. Moran. 2009. Arsenophonus, an emerging clade of intracellular symbionts with a broad host distribution. *BMC Microbiol.* 9:143.

Oliver, K.M., P.H. Degnan, M.S. Hunter, and N.A. Moran. 2009. Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science*. 325:992-994.

Oliver, K.M., P.H. Degnan, G.R. Burke, and N.A. Moran. 2010. Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annu. Rev. Entomol.* 55:247-266.

Pais, R., C. Lohs, Y. Wu, J. Wang, and S. Aksoy. 2008. The obligate mutualist Wigglesworthia glossinidia influences reproduction, digestion, and immunity processes of its host, the tsetse fly. *Appl. Environ. Microbiol.* 74:5965-5974.

Pontes, M.H., and C. Dale. 2011. Lambda Red-mediated genetic modification of the insect endosymbiont Sodalis glossinidius. *Appl. Environ. Microbiol.* 77:1918-1920.

Pontes, M.H., K.L. Smith, L. De Vooght, J. Van Den Abbeele, and C. Dale. 2011.

Attenuation of the sensing capabilities of PhoQ in transition to obligate insect-bacterial association. *PLoS Genet*. 7:e1002349.

Russell, J. A., Latorre, A., Sabater-Munoz, B., Moya, A., and N.A. Moran. 2003. Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. *Mol. Ecol.*, 12:1061-1075.

Russell, J.A., and N.A. Moran. 2006. Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. *Proc. Biol. Sci*. 273:603-610.

Sabri, A., P. Leroy, E. Haubruge, T. Hance, I. Frere, J. Destain, and P. Thonart. 2010. Isolation, pure culture and characterization of Serratia symbiotica sp. nov., the R-type of secondary endosymbiont of the black bean aphid Aphis fabae. *Int. J. Syst. Evol. Microbiol*. 61:2081-2088.

Scarborough, C.L., Ferrari, J., and H.C.J. Godfray. 2005. Aphid protected from pathogen by endosymbiont. *Science*. 310:1781.

Thao, M.L., and P. Baumann. 2004. Evidence for multiple acquisition of Arsenophonus by whitefly species (Sternorrhyncha: Aleyrodidae). *Curr. Microbiol*. 48:140-144.

Trowbridge, R.E., K. Dittmar, and M.F. Whiting. 2006. Identification and phylogenetic analysis of Arsenophonus- and Photorhabdus-type bacteria from adult Hippoboscidae and Streblidae (Hippoboscoidea). *J. Invertebr. Pathol*. 91:64-68.

Tsuchida, T., R. Koga, and T. Fukatsu. 2004. Host plant specialization governed by facultative symbiont. *Science*. 303:1989.

Wilkes, T.E., Duron, O., Darby, A.C., Hypsa, V., Novakova, E., and G.D.D. Hurst. 2011. The genus Arsenophonus. In Manipulative Tenants: Bacteria Associated with Arthropods. E. Zchori-Fein, and K. Bourtzis, editors. CRC Press, Boca Raton, FL. 225-244.

Wilkinson, T.L., T. Fukatsu, and H. Ishikawa. 2003. Transmission of symbiotic bacteria Buchnera to parthenogenetic embryos in the aphid Acyrthosiphon pisum (Hemiptera: Aphidoidea). *Arthropod Struct. Dev*. 32:241-245.

CHAPTER 2

CHARACTERISTICS OF THE COMPLETE GENOME SEQUENCE OF CANDIDATUS ARSENOPHONUS ARTHROPODICUS

Abstract

Hippoboscid louse flies are blood-feeding parasites of pigeons that harbor an ancient obligate symbiont required for nutritional supplementation, as well as a more recently derived symbiont, *Candidatus* Arsenophonus arthropodicus. This recent symbiont is a member of the *Arsenophonus* clade, which consists of a wide array of closely related symbionts that are associated with many distantly related insect hosts. Symbionts in this clade are phylogenetically related to the free-living/opportunistic human pathogen *Proteus mirabilis* and the nematode symbiont/insect pathogen, *Photorhabdus luminescens*. We have sequenced the genome of *Ca.* A. arthropodicus and identified factors shared with other insect symbionts as well as virulence factors related to those used by bacterial pathogens to interact with hosts. The genome of *Ca.* A. athropodicus shows a bacterium in transition from a free-living lifestyle to a permanent insect association. The presence of virulence factors and toxin-encoding genes and gene fragments suggests that *Ca.* A. athropodicus has evolved from a pathogenic ancestor to a mutualistic insect symbiont through attenuation of pathogenic interactions that would have detrimental host effects and a transition from horizontal to vertical transmission.

Introduction

Many insects have developed intimate relationships with mutualistic bacterial symbionts that allow them to exploit a novel niche, and these interactions have likely played an important role in shaping the evolution of insect species (Steinert et al., 2000). While some of these relationships are known to have a recent origin of association, others are known to be ancient and obligate in nature. It has been predicted that these mutualists evolved from ancestors that were insect parasites or other parasites that used insects as vectors, and that they transitioned through a switch in bacterial transmission strategy from horizontal to vertical, concomitant with an attenuation in parasite virulence, which benefits the host and increases the likelihood of symbiont transmission (Ewald, 1987; Steinert et al., 2000; Weeks et al., 2007).

Genome sequences for many bacterial symbionts of insects have been generated in the last decade and have provided a framework for understanding the genomic changes that occur over time as bacteria take up residence in their insect hosts. Over the course of long-term host restriction, bacterial symbionts undergo substantial gene inactivation and loss due to the relaxation of selection on genes that are not essential in the symbiotic lifestyle (McCutcheon and Moran, 2012). Genome degeneration is proposed to be exacerbated by an increased rate of fixation of slightly deleterious mutations which occurs as a consequence of frequent population bottlenecks during host reproduction (Moran, 1996; Moya et al., 2008). As might be expected, recently derived symbionts are found to have larger gene inventories than ancient symbionts, and the recently derived symbionts maintain genes that share sequence homology with virulence factors used by related bacterial pathogens (Dale et al., 2001; Degnan et al., 2009; Moya et al., 2008).

For example, the tsetse fly symbiont, *Sodalis glossinidius*, is known to utilize type III secretion systems to facilitate insect cell invasion and intracellular proliferation (Dale et al., 2001; Dale et al., 2005). These secretion systems are known to play major roles in pathogen interaction with host cells, allowing pathogens to deliver effector proteins directly into the cytoplasm of target cells (Hueck, 1998). However, these virulence factors are not found in the substantially reduced genomes typical of ancient symbionts, most likely due to the fact that they have been lost in the transition to a highly specialized, obligate association (Dale and Moran, 2006). Therefore, some virulence factors that are required for the lifestyle of bacterial pathogens may play only a temporary role in the transition from parasitism to mutualism.

Here we present the genome sequence of *Candidatus* Arsenophonus arthropodicus, a recently derived symbiont of the pigeon louse fly, *Pseudolynchia canariensis*. Members of the *Arsenophonus* group of bacteria are found in a broad range of insect hosts which are phylogenetically diverse (Dale et al., 2006; Novakova et al., 2009) and recent screens identified *Arsenophonus* species in 5% of the insects surveyed (Duron et al., 2008). The interactions between *Arsenophonus* symbionts and their insect hosts are quite variable, ranging from beneficial mutualisms to reproductive parasitisms, and involve mechanisms of both vertical and horizontal transmission (Wilkes et al., 2011). The genome of *Ca.* A. arthropodicus shares traits common to other recently derived insect symbionts, including a genome size and gene inventory that is reduced from free-living relatives but substantially larger than those of ancient symbionts. The presence of genes or gene fragments that share sequence homology with known virulence factors involved in pathogenesis suggests that the *Ca.* A. arthropodicus - louse fly

symbiosis originated from a bacterial ancestor that was an insect pathogen or other pathogen that utilized insects, such as louse flies, as a transmission vector.

Materials and Methods

DNA preparation and library construction for Sanger sequencing

Ca. A. arthropodicus was isolated from louse fly pupae and liquid cultures were maintained at 28°C in Mitsuhashi and Maramorosch medium (MM medium) as described previously (Dale et al., 2006). Purified genomic DNA was extracted using a DNeasy Blood and Tissue Kit (Qiagen) from 3 mls of *Ca.* A. arthropodicus culture harvested by centrifugation in mid-log phase of growth. The genomic DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice to an average size range of 8-12 kb. Sheared DNA was blunt end-repaired with T4 DNA polymerase and phosphorylated with T4 polynucleotide kinase prior to blunt-end ligation with biotinylated adaptor oligonucleotides. Adaptored DNA was then purified by capture on streptavidin coated beads. Plasmid vector pWD42 (Robb et al., 2001), a copy-number inducible derivative of plasmid R1, was prepared and ligated with adaptors complementary to the insert adaptors and the insert DNA was then annealed to the adaptored vector. Vectors containing genomic DNA inserts were transformed into chemically competent *E. coli* XL-10 Gold cells (Stratagene) and plated on Terrific Broth (TB) media plates with ampicillin to select for transformants.

Plasmid induction and Sanger sequencing

After overnight growth at 30°C, clones were picked into liquid TB and grown for 16 hours at 30°C. Runaway plasmid replication was induced by incubating in a 42°C

shaking water bath for 2.25 hours. Standard alkaline lysis plasmid preparation procedures were used to collect plasmid DNA and plasmids were digested with NotI (NEB) to check insertion rate frequency. Paired end sequencing reads were generated from the inserts using a BigDye terminator cycle sequencing kit (Applied Biosystems) with primers complementary to vector sequences flanking the insert. DNA was ethanol precipitated to remove excess fluorescent terminators and analyzed on an ABI 3730 96-capillary instrument. Paired-end Sanger sequencing reads were then analyzed and assembled using the Phred/Phrap/Consed (Ewing and Green, 1998; Ewing et al., 1998; Gordon et al., 1998) assembly software with default parameters.

Library construction and Illumina sequencing

A DNA sequencing library was constructed using the Illumina Paired-End DNA Sample Prep Kit according to manufacturer's instructions. All DNA purification steps were performed using a Qiagen PCR purification kit. Briefly, 5 µg of purified *Ca.* A. arthropodicus genomic DNA was fragmented by nebulization. Fragmented DNA was end-repaired with T4 DNA polymerase and DNA polymerase I Klenow fragment. Single 'A' bases were added to the 3' ends of the DNA fragments with Klenow exo (3' to 5' exo minus) and purified fragments were ligated to adaptors with 3' 'T' overhangs. Ligated products were subjected to gel electrophoresis and products in the 150-200 bp range were gel extracted and purified. DNA fragments with adaptors on both ends were enriched by 18 cycles of PCR using primers complementary to sequences on the adaptor ends. Cluster generation and 36-bp paired-end sequencing was performed on an Illumina Genome Analyzer II.

Hybrid assembly of sequencing reads

A hybrid assembly of paired-end Sanger and Illumina reads was generated using the CLC Genomics Workbench (CLC bio, Aarhus, Denmark) high-throughput sequencing *de novo* assembly algorithm. Contigs were manually inspected for misassemblies and edited using Consed (Gordon et al., 1998). Sequence gap closures were performed by primer walking on the appropriate gap spanning Sanger library clones. Remaining physical gaps were closed by Sanger sequencing of PCR products spanning gaps, which were generated by touchdown PCR using Phusion DNA Polymerase (Finnzymes). The assembly was verified by aligning all sequencing reads back to the assembly consensus sequence to confirm that an equal coverage distribution was obtained and that paired-end sequences were correctly arranged.

Genome annotation and analysis

The genome was annotated using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP), which uses a combination of GeneMark and Glimmer to predict genes (Borodovsky and McIninch, 1993; Lukashin and Borodovsky, 1998; Delcher et al., 1998), and uses tRNAscan-SE (Lowe and Eddy, 1997) to predict tRNAs. Ribosomal RNAs were predicted using BLAST searches (Altschul et al., 1990) against an RNA database. The Conserved Domain Database (Marchler-Bauer et al., 2011) and Cluster of Orthologous Groups (Tatusov et al., 1997) were used to assign putative functions to genes. The annotation obtained from the PGAAP was then manually examined in Artemis (Rutherford et al., 2000) and each predicted CDS was verified by performing a BLAST search against the NCBI non-redundant protein database to confirm start and stop codons and to identify candidate pseudogenes containing frameshifting

mutations or modified start/stop codons. Circular genome representations were generated using DNAPlotter (Carver et al., 2009). Sequence alignments were performed using Clustal X (Larkin et al., 2007). Metabolic capability was assessed using Pathway Tools (Dale et al., 2010) and the KEGG (Kyoto Encyclopedia of Genes and Genomes; Kanehisa and Goto, 2000) automatic annotation server (KAAS; Moriya et al., 2007).

Phylogenetics

A bacterial species phylogeny was generated based on the concatenated nucleotide sequences of seven conserved genes (*frr, gyrB, infB, pheS, prfA, pth,* and *tsf*). The concatenated gene sequences were aligned using Clustal X (Larkin et al., 2007) and maximum likelihood trees were constructed using PhyML (Guindon and Gascuel, 2003). PhyML parameters were: HKY85 model of sequence evolution, starting from 20 random trees, with the proportion of invariable sites estimated from the data and 100 bootstrap replicates.

Results and Discussion

Genome sequencing

The *Ca.* A. arthropodicus genome sequence was generated via a hybrid sequencing approach using Sanger sequencing and next-generation Illumina sequencing of paired-end libraries. The Sanger and Illumina sequencing reads were assembled together using the CLC Genomics Workbench *de novo* assembly algorithm and yielded 123 contigs ranging in size from 95 bp to 187,172 bp (Table 2.1). Sequence gaps in the assembly were closed by primer walking gap spanning clones, while physical gaps were closed by primer walking PCR products (Table 2.1). The final assembly comprises three

Table 2.1. Statistics of *de novo* hybrid assembly of Illumina and Sanger reads.

Chromosomal contigs in initial assembly (size)	119 (2,726,065 bp)
Average chromosomal contig length	22,908 bp
Plasmid contigs in initial assembly (size)	4 (95,856 bp)
Average genome coverage	259X
Number of reads in assembly	25,920,686
Number of primer walks	386
Chromosomal contigs in final assembly (size)	3 (2,873,045 bp)
Plasmids closed in final assembly (total size)	3 (98,909 bp)

large contigs (1,437,952 bp, 945,492 bp, and 489,601 bp) that make up the chromosome, separated by gaps of approximately 20 kb, and three closed circular plasmid contigs. Sequence reads that are not integrated into these contigs assemble into small contigs (< 2 kb) of repetitive phage sequences that have not yet been placed into the chromosomal gaps. Given the coverage obtained by both sequencing methods (~3X Sanger coverage, ~200X Illumina coverage), it is highly unlikely that there are any remaining unique coding sequences within these gaps.

General features of the genome

The genome features of *Ca.* A. arthropodicus were compared with other related gamma-proteobacteria for which genome sequences are available, that have lifestyles ranging from free-living bacteria to ancient, obligate symbionts of insects (Table 2.2). A phylogeny of these bacteria was constructed (Fig. 2.1) based on concatenated nucleotide sequence alignments of seven conserved orthologous genes using PhyML (Guindon and Gascuel, 2003). Based on this phylogeny, and supported by other phylogenetic constructions based on 16S ribosomal RNA sequences (Dale et al., 2006; Darby et al., 2010; Novakova et al., 2009), the closest sequenced relatives to *Ca.* A. arthropodicus are the free-living *Proteus mirabilis* and the nematode symbiont/insect pathogen

Table 2.2. Genome features of bacteria with different lifestyles.

	Free-living or symbiotic relatives				**Recently acquired symbionts**			**Ancient symbionts**	
	Escherichia coli K12	*Dickeya dadantii*	*Proteus mirabilis*	*Photorhabdus luminescens*	*Ca.* A. arthropodicus	*Sodalis glossinidius*	*Arsenophonus nasoniae*	*Buchnera aphidicola*	*Wigglesworthia glossinidia*
Chromosome (bp)	4,639,221	4,922,802	4,063,606	5,688,987	2,873,045	4,171,146	3,567,128[a]	640,681	697,742
GC content (%)	50.8	56.3	38.9	42.8	38.1	54.7	37.3	26.2	22.5
Predicted CDS	4,146	4,549	3,693	4,683	2,231	2,432	3,332	571	611
Pseudogenes	179	22	24	222	395	1,501	135	13	8
rRNA operons	7	7	7	7	7	7	8-10[b]	2	2
tRNA	89	75	83	85	64	69	52	32	34
Plasmids	0	0	1	0	3	3	2 or more[b]	2	1

[a] Size of chromosome and plasmid scaffolds.
[b] Estimates based on fractured assembly.

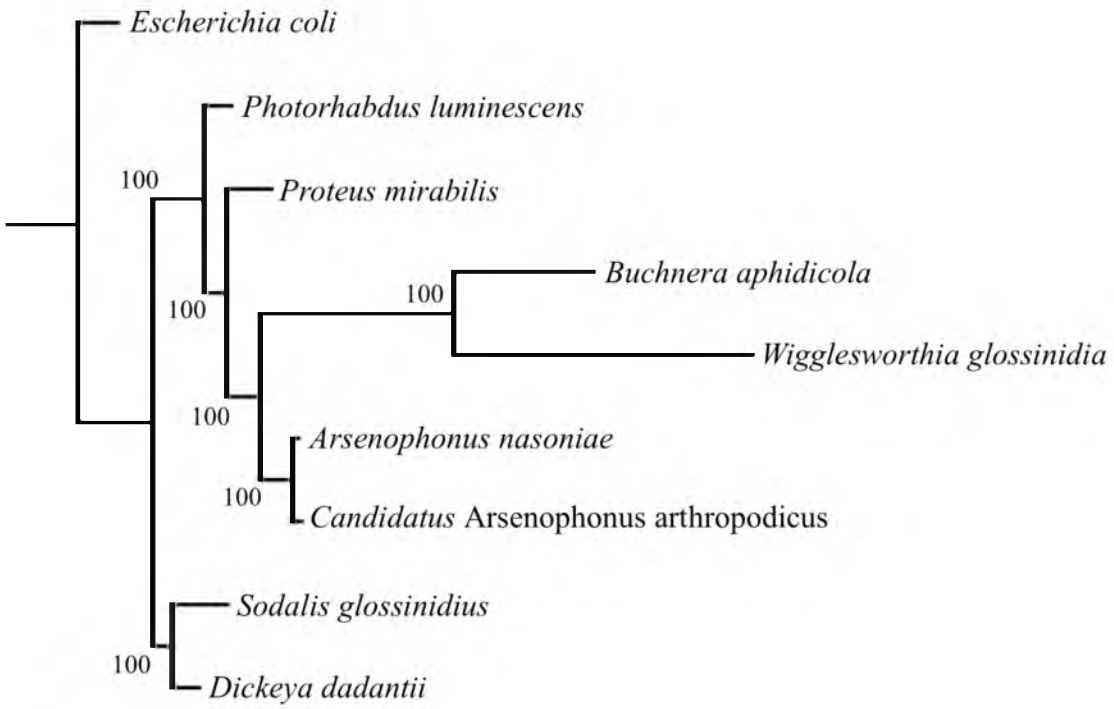


Figure 2.1. Phylogenetic position of *Ca.* A. arthropodicus based on concatenated sequences of seven conserved orthologous genes. Phylogenetic analysis based on maximum likelihood estimation using PhyML (Guindon and Gascuel, 2003), with bootstrap values indicated, based on 100 replicates. GenBank accession numbers for sequences used for tree construction are provided in Table 2.2.

Photorhabdus luminescens (Fig. 2.2). Trees were also generated with nhPhyML (Boussau and Guoy, 2006), using a nonhomogenous model of sequence evolution that allows for varying rates of sequence evolution on different branches of the tree, and the resultant tree topology remained the same.

The *Ca.* A. arthropodicus genome is predicted to comprise three plasmids (Fig. 2.2) and a single 2.88 Mb chromosome (Fig. 2.3), which is reduced in size from that of its close free-living relative, *Proteus mirabilis*, that has a genome size of 4.06 Mb (Table 2.2; Pearson et al., 2008). The genome of *Ca.* A. arthropodicus displays a moderate bias toward AT, with an increased AT content of 61.9%, a feature which is not typical of free-

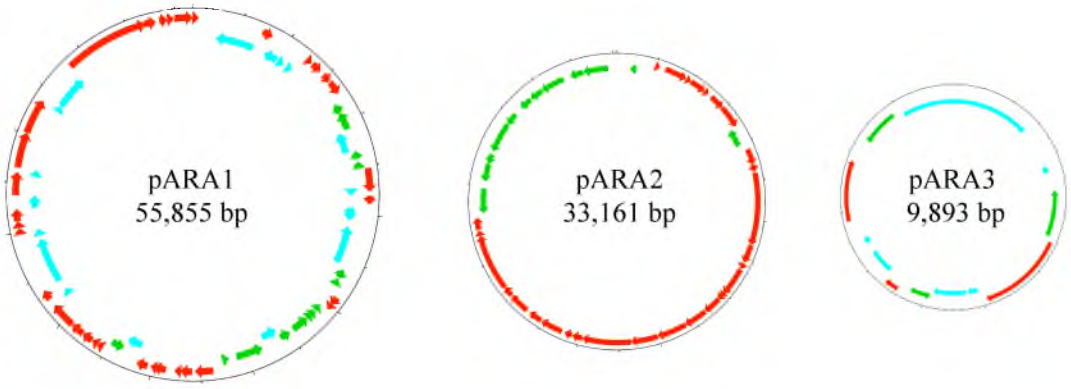


Figure 2.2. Plasmids of *Ca.* A. arthropodicus. Depictions of the three plasmids present in the *Ca.* A. arthropodicus genome. Red arrows indicate genes on leading strand, green arrows indicate genes on lagging strand, blue arrows indicate pseudogenes.

living bacteria and recently-associated insect symbionts, such as *Escherichia coli* (Blattner et al., 1997), *Dickeya dadantii* (Glasner et al., 2011) and *Sodalis glossinidius* (Toh et al., 2006), but which is often observed in more ancient insect symbionts, such as *Buchnera aphidicola* (Shigenobu et al., 2000) and *Wigglesworthia glossinidia* (Akman et al., 2002). However, the genome sequences of *Arsenophonus* relatives *P. mirabilis* (Pearson et al., 2008) and *P. luminescens* (Duchaud et al., 2003) are also biased, with AT content around 60% (Table 2.2). Thus, a bias toward AT in *Ca.* A. arthropodicus may be a feature that is inherent to it and its close relatives rather than a result of insect host restriction.

The *Ca.* A. arthropodicus genome contains numerous pseudogenes (~15% of the predicted genes; Fig. 2.3), a feature which is common to other symbionts that are in the early stages of genome degeneration, such as *S. glossinidius,* in which ~38% of predicted genes are anticipated to be pseudogenes (Belda et al., 2010; Toh et al., 2006), but which is not a prominent feature of free-living bacteria and long established endosymbionts,

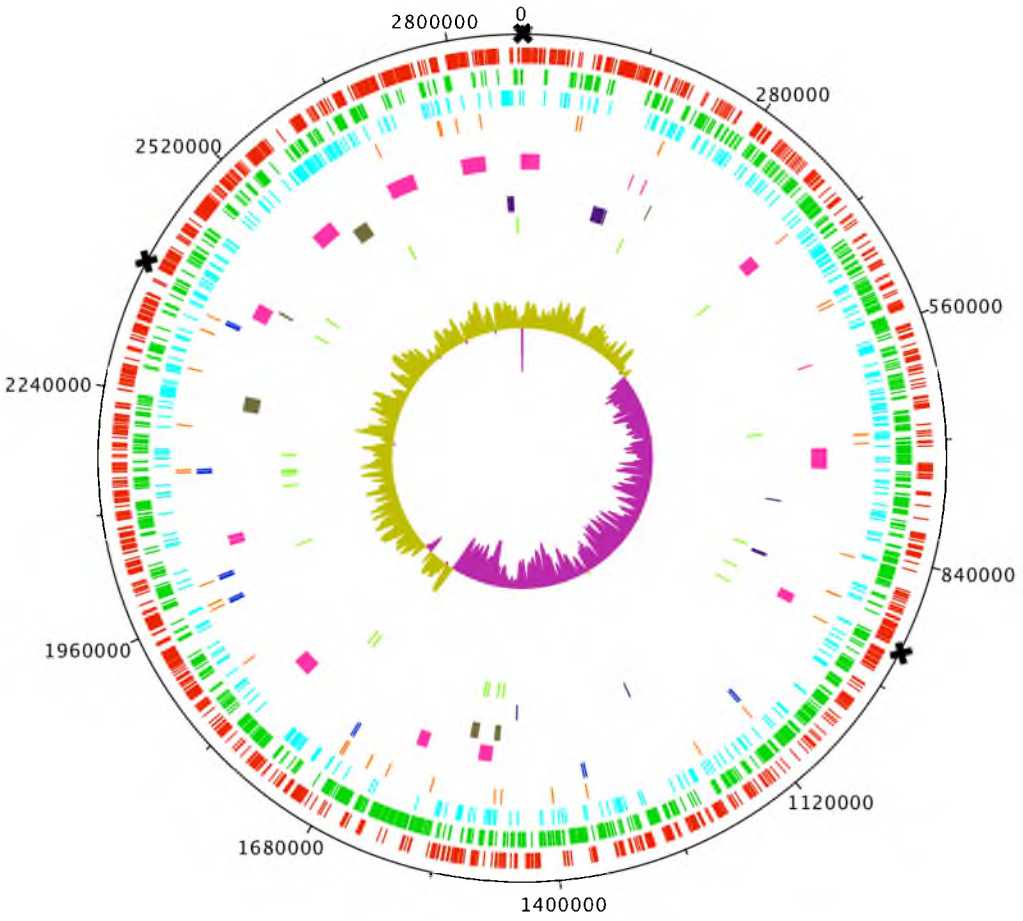


Figure 2.3. Genome features of *Ca.* A. arthropodicus. From outer track: genes encoded on the leading strand (red), genes on lagging strand (green), pseudogenes (light blue), tRNAs (orange), rRNAs (dark blue), phage genes (pink), type III secretion system genes (brown), genes encoding predicted toxins- TC, RTX, Mcf (purple), ICS/IS911 repeats (light green). Innermost circle depicts GC skew. Gaps between contigs in the scaffold are indicated by (✖) and contigs were ordered based on PCR analysis.

both of which tend to maintain relatively low proportions of pseudogenes (Table 2.2). Additionally, the genome of *Ca.* A. arthropodicus contains repetitive insertion sequences and phage regions (15% of the CDSs; Fig. 2.3), another characteristic that is common to recently acquired symbionts such as *S. glossinidius* (21% of CDSs; Belda et al., 2010; Toh et al., 2006) and the aphid symbiont, *Hamiltonella defensa* (21%; Degnan et al., 2009), but occur more infrequently in free-living bacteria and ancient symbionts (McCutcheon and Moran, 2012).

To better understand the gene inventory of *Ca.* A. arthropodicus, Cluster of Orthologous Groups (COG; Tatusov et al., 1997) information was assigned to predicted genes (Fig. 2.4). There are 698 genes (31% of the CDSs) that are currently annotated as encoding "hypothetical proteins" and could not yet be assigned to COG categories. For the remaining gene inventory, the COG categories that have accumulated the lowest percentages of pseudogenes are F (3.1% - genes involved in nucleotide transport and metabolism), O (4.4% - Posttranslational modification, protein turnover and chaperones) and J (4.8% - Translation, ribosomal structure and biogenesis). The increased retention of genes in these categories is predicted to be due to the fact that they provide core functions involved with bacterial survival and replication within the insect host. COG categories showing increased percentages of pseudogenes include categories U (36.8% - Intracellular trafficking, secretion and vesicular transport), N (16.5% - Cell motility) and T (15.4% - Signal transduction mechanisms). A higher proportion of genes within these categories may be unneccessary or redundant within the static insect environment, encoding products that might be dispensible or are provided by the insect host. For example, COG category U includes genes involved in protein secretion systems and

conjugation, such as the type III and type IV secretion systems, which are often absent in ancient symbionts, including *W. glossinidia* and *B. aphidicola* (Akman et al., 2002; Shigenobu et al., 2000). Additionally, many ancient symbionts, such as *Blochmannia* and *Baumannia* species have lost genes associated with flagellar motility (COG category N) over the course of their intimate intracellular relationship with an insect host (Toft and Fares, 2008), indicating that motility is not a requirement for insect symbionts. However, some ancient intracellular symbionts, such as *W. glossinidia* and *B. aphidicola,* have retained a subset or near complete inventory of the flagellar genes, which may be required during specific life stages or may function as a protein secretion system (Akman et al., 2002; Maezawa et al., 2006). The necessity for a number of regulatory signal transduction systems (COG category T) may be reduced in the constant environment within the body of an insect, leading to a loss of a higher proportion of genes involved in altering gene expression in response to changing environmental conditions. Preferential loss of genes in these COG categories suggests that many systems of regulation, secretion and motility may not be required once a bacterium becomes sequestered in a static, protected host environment. Therefore, these genes are under relaxed selection to be maintained and can more easily be inactivated and lost from the genome, without detrimental fitness effects.

Plasmids

Ca. A. arthropodicus harbors three plasmids of 55,855 bp, 33,161 bp and 9,893 bp (Fig. 2.2). The largest plasmid, pARA1, maintains open reading frames (ORFs) that share homology with *tra* conjugative transfer genes, although many have been truncated or have accumulated frameshifting mutations, which is predicted to render them

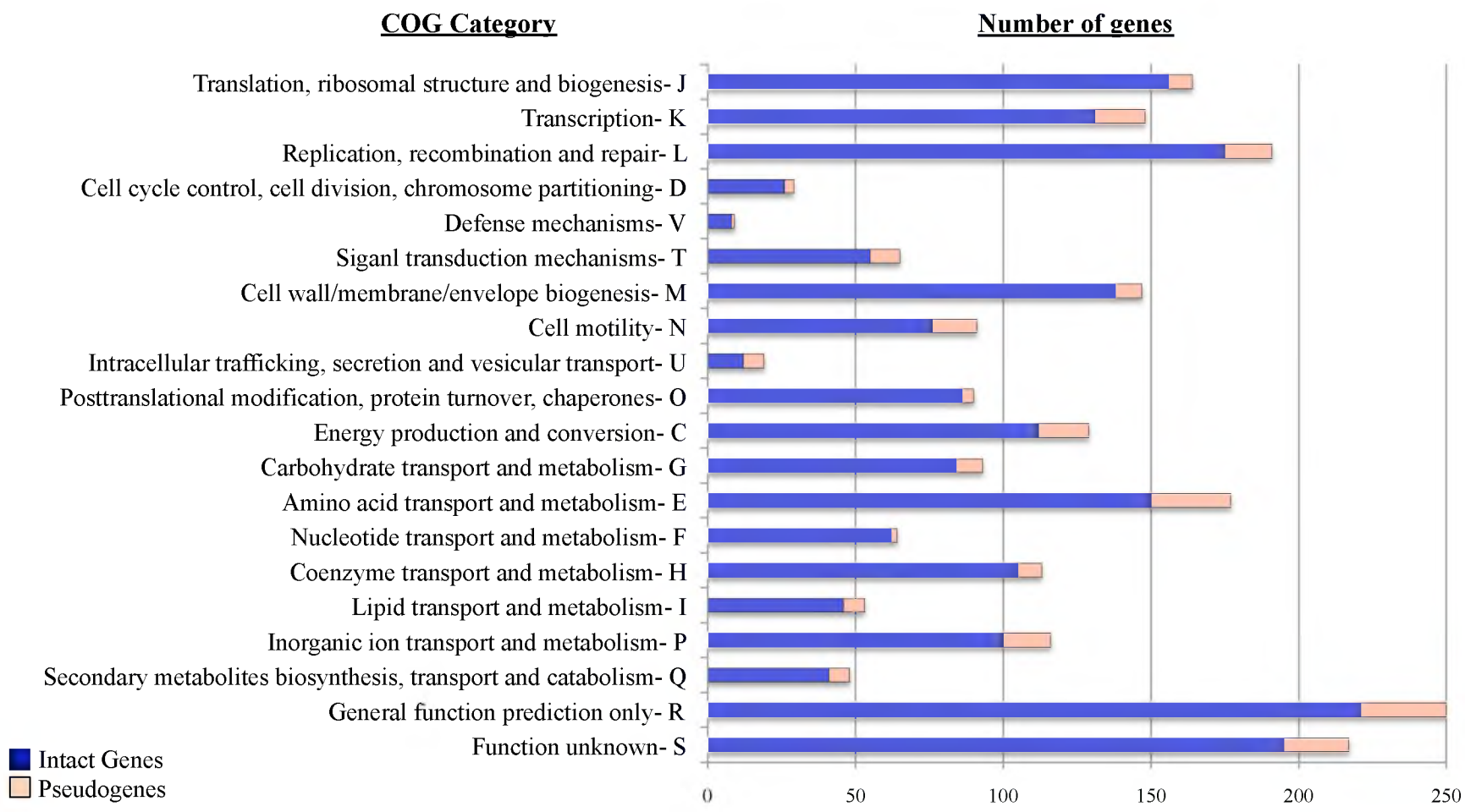


Figure 2.4. COG category classification of genes. Dark bars indicate the number of intact predicted coding sequences and light bars indicate the number of pseudogenes per category.

nonfunctional. The loss of functionality of these genes is not surprising given that this bacterium likely has little opportunity to engage in parasexual recombination since it is a symbiont that is sequestered in an insect host (McCutcheon and Moran, 2012). Intact coding sequences (CDSs) carried by pARA1 include a set of ascorbate-specific phosphotransferase system (PTS) components. There are additional PTS components specific to other sugars present in the chromosome, however, this system might play an important metabolic role, given that plasmid-borne genes may have increased expression relative to the chromosomal genes since plasmids are often present in multiple copies, or may provide greater regulatory control over gene expression (Moran et al., 2003). The presence of these plasmid-borne PTS genes may allow *Ca.* A. arthropodicus to utilize ascorbate as an additional carbohydrate source. Since ascorbate is found in high levels in white blood cells (Corti et al., 2010), this vitamin may be available in the louse fly's blood meal for use by *Ca.* A. arthropodicus.

Plasmid pARA2 encodes a set of genes that share homology with the *tri/virB* type IV secretion system genes, which are found in a variety of bacteria and are also widespread throughout the *Wolbachia,* possibly playing a role in their associations with insect hosts (Pichon et al., 2009). The *vir* genes present on pARA2 all appear to be intact, suggesting that these genes are important in the symbiosis or that this plasmid may be a more recent acquisition that has not yet undergone any degenerative evolution. Type IV secretion systems commonly function in conjugation and DNA transport as well as transport of effector molecules into the cytoplasm of eukaryotic cells (Rances et al., 2008).

A CDS homologous to the gene encoding the queuosine biosynthesis protein, QueC, is present on both the pARA1 and pARA2 plasmids, with a copy also located on the *Ca.* A. arthropodicus chromosome. Queuosine is involved in tRNA modifications that are predicted to enhance the efficiency of protein synthesis (Cicmil and Huang, 2008). Analysis of the three *queC* CDSs shows that the two plasmid-borne copies are 94% identical at the amino acid level but only share 70% identity with the chromosomal *queC*. The chromosomally-encoded QueC shares 97% amino acid sequence identity with the QueC of *Arsenophonus nasoniae*, but has an N-terminal truncation of 55 amino acyl residues. It is unclear whether this truncation may have diminished or abolished the function of QueC, necessitating the need for recruitment of additional gene copies, or whether the acquisition of additional plasmid-borne copies led to relaxed selection on the chromosomal copy, making its maintenance unneccessary. Either way, QueC is likely important in the lifestyle of *Ca.* A. arthropodicus, given that it is present on multiple plasmids that may result in enhanced expression levels.

Plasmid pARA3 is a small plasmid composed mostly of pseudogenes, however, it maintains intact genes encoding a LuxR-like transcriptional regulator, a colicin Ib immunity protein and a cytosine permease. There are two additional genes encoding cytosine permeases located within the chromosome of *Ca.* A. arthropodicus. Multiple copies of cytosine permeases are also present in the genomes of the insect symbionts *S. glossinidius* and *A. nasoniae* (Toh et al., 2006; Darby et al. 2010), suggesting an important role for cytosine transport in these recently established symbionts.

Genes encoding structural RNAs

Candidatus Arsenophonus arthropodicus maintains 64 tRNAs comprising all 20 amino acids and seven ribosomal RNA operons consisting of three different organizational structures (Fig. 2.5A). The rRNA operons differ in the identities of interspersing tRNAs as well as in the number of 5S subunits included, a variation which is also observed in the close free-living relative of *Ca.* A. arthropodicus, *P. mirabilis* (Pearson et al., 2008). Since bacterial phylogenies often utilize rRNA sequences which can be affected by heterogeneity in the sequences (Sorfova et al., 2008) we aligned the paralogous copies of these rRNA genes to determine if they encoded divergent sequences. The 5S subunits were relatively polymorphic (Fig. 2.5B), while the 16S subunit showed little heterogeneity (Fig. 2.5C) and the 23S subunit was completely conserved in all seven operons.

Metabolic capability

Analysis using the KEGG automatic annotation server (KAAS; Moriya et al., 2007) indicates that *Ca.* A. arthropodicus retains complete metabolic pathways for glycolysis, gluconeogenesis, the tricarboxylic acid cycle and the pentose phosphate pathway. Additionally, *Ca.* A. arthropodicus retains pathways for the synthesis of certain B vitamins, including biotin, thiamine, nicotinic acid and pantothenate. B vitamins are predicted to be lacking in the louse fly host's blood diet and may need to be supplemented by bacterial symbionts, in a similar manner to that observed in tsetse flies. Both the primary and secondary symbionts of the tsetse fly, *Wigglesworthia glossinidia* and *S. glossinidius*, respectively, are predicted to play a role in diet supplementation, providing essential B vitamins to the tsetse fly that are not present in sufficient quantities

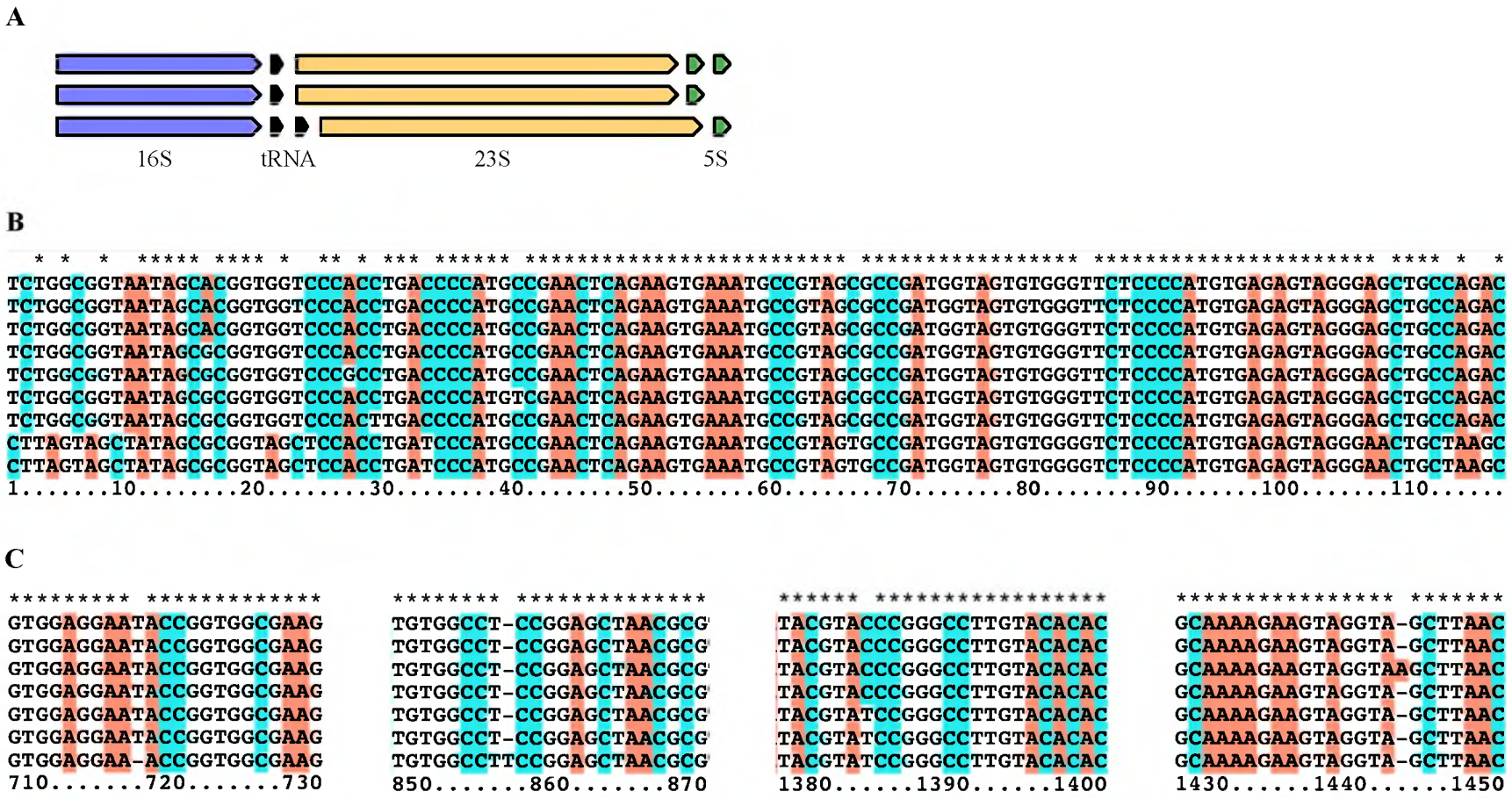


Figure 2.5. Ribosomal RNA operon organization and heterogeneity. A. Three different rRNA operon structures in *Ca.* A. arthropodicus. B. Alignment of complete 5S rDNA sequences from all seven operons. C. Portions of 16S sequence alignments containing non-conserved positions. Asterisks indicate sites conserved in all sequences. Base numbers are indicated below alignment.

in vertebrate blood (Akman et al., 2002; Nogge, 1981). *Ca.* A. arthropodicus may play an analogous role to *S. glossinidius* in the louse fly, which incidentally also maintains an ancient bacteriome-associated symbiont that is closely related to *W. glossinidia* (Dale et al., 2006).

Motility

Many recently acquired and some ancient insect symbionts retain genes necessary to assemble functional flagella for locomotion (Akman et al., 2002; Darby et al., 2010; Degnan et al., 2010; Toh et al., 2006). Due to their prevalence in many symbiont genomes, it has been postulated that flagella may play an important role in allowing symbionts to infect specific tissues in the host insect or may be required for motility during certain insect life stages, such as during insect reproduction when symbionts are maternally transmitted (Rio et al., 2012). Flagellar genes are present in *Ca.* A. arthropodicus and are located within a single 38 kb island. However, it appears that genes encoding the essential flagellar components FliF, FliI, FliJ, FliP, FliS, and FlhA (Berg, 2003) have been truncated by frameshifting mutations or premature stop codons in *Ca.* A. arthropodicus (Table 2.3). Additionally, genes encoding truncated components of the FliG and FliH proteins have been fused into a single ORF. We tested for *Ca.* A. arthropodicus motility using swarm plate assays and motility was not observed. Furthermore, neither basal bodies nor complete flagella were visualized using transmission electron microscopy (data not shown). Based on this information, we predict that *Ca.* A. arthropodicus do not produce functional flagella. However, it is possible that a limited inventory of flagellar gene components provides an alternative

Table 2.3. Flagellar components in *Ca.* A. arthropodicus.

	Status[a]	Length[b]	% Identity[c]	Ortholog length[d]	Function
FliZ	+	171	76	176	Regulation
FliA	+	240	73	240	Sigma factor
FliC	+	378	58	363	Flagellin
FliD	+	468	47	472	Filament cap
FliS	p	47	58	132	Chaperone (FliC)
FliT	+	112	40	117	Chaperone (FliD)
FliE	p	42	72	110	Basal body
FliF	p	417	54	573	M ring
FliG	p[e]	347[e]	72	332	Rotor/switch complex
FliH	p[e]	347[e]	57	240	Protein export
FliI	p	289	71	457	Export ATPase
FliJ	p	108	49	148	Chaperone
FliK	p	322	37	470	Hook-length control
FliL	+	160	60	160	Rotational control
FliM	+	342	65	343	Switch complex
FliN	+	139	71	136	Switch complex
FliO	+	148	35	148	Protein export
FliP	p	146	80	256	Protein export
FliQ	+	89	62	89	Protein export
FliR	+	267	57	260	Protein export
FlgL	+	312	34	314	Hook-filament junction
FlgK	+	548	39	547	Hook-filament junction
FlgJ	p	65	54	328	Peptidoglycan hydrolase
FlgI	p	197	73	368	P-ring
FlgH	+	203	74	247	L-ring
FlgG	+	260	81	260	Distal rod
FlgF	p	134	60	251	Proximal rod
FlgE	+	419	57	406	Hook
FlgD	+	238	45	264	Hook assembly
FlgC	+	134	78	134	Proximal rod
FlgB	+	135	64	137	Proximal rod
FlgA	p	107	45	218	P-ring assembly
FlgM	+	100	33	99	Anti-sigma factor
FlgN	p	104	45	146	Chaperone (FlgK/FlgL)
FlhA	p	422	80	696	Protein export
FlhB	+	383	64	382	Protein export
MotB	+	293	59	349	Stator
MotA	+	297	71	297	Stator
FlhC	+	189	76	193	Regulation
FlhD	+	125	80	116	Regulation

[a] (+) indicates intact component, (p) indicates a predicted pseudogene.
[b] Length of largest component ORF in amino acyl residues.
[c] Percent amino acid sequence identity shared by *Ca.* A. arthropodicus and *P. mirabilis* orthologs.
[d] Length of *P. mirabilis* orthologs in amino acyl residues.
[e] Truncated FliG/FliH components fused into single ORF of indicated size.

function, perhaps in terms of protein secretion, as has been predicted for *B. aphidicola* (Maezawa et al., 2006).

Insecticidal toxins

The *Ca.* A. arthropodicus genome retains remnants of the high molecular weight insecticidal toxin complexes (TCs) discovered in *P. luminescens* (Bowen et al., 1998). *Photorhabdus luminescens* is a mutualistic symbiont of entomopathogenic nematode hosts and plays a vital role in its host life cycle, due to the large number of toxins it produces that are targeted toward insects (Waterfield et al., 2009). When a nematode host harboring *P. luminescens* infects an insect, it releases the bacterial symbionts into the insect hemocoel. Under the conditions encountered in an insect host, *P. luminescens* expresses the TCs and other toxic compounds that rapidly kill the insect and both nematode and symbiont then obtain nutrition from and reproduce within the cadaver (Waterfield et al., 2009). The TC products encoded by *P. luminescens* have been shown to have high levels of toxicity toward a wide range of insect hosts (Blackburn et al., 2005; Waterfield et al., 2009). Therefore, these insecticidal toxin genes would not be expected to be present in the genome of a mutualistic symbiont of insects.

For complete toxicity of the TCs, multiple toxin components are required, of three different types: a TcA-like component, a TcB-like component, and a TcC-like component (ffrench-Constant and Waterfield, 2005). Located adjacent to each other in the *Ca.* A. arthropodicus genome are two genes, *tcdA* and *tcdB*, which are homologs of members of the TcA and TcB classes, respectively. However, both ORFs contain one or more frameshifting mutations, likely rendering them nonfunctional. There are two TcC component-encoding CDSs sharing homology with *tccC* genes of *P. luminescens*, located

in separate regions of the genome. Both *tccC* CDSs maintain intact reading frames of a similar size to homologs in *P. luminescens* and other related bacteria (Table 2.4), even though their predicted partner components have acquired frameshifting mutations.

The presence of intact insecticidal *tccC* gene homologs in Ca. A. arthropodicus is intriguing, given that the additional genic components required to produce a functional toxin complex have been inactivated by mutations. The TccC components in *P. luminescens* show sequence similarity to other RHS/YD-repeat proteins and have been shown to have ADP-ribosyltransferase activity and function in actin polymerization and cytoskeletal rearrangement of insect cells (Lang et al., 2010a; 2010b). This is also the mode of action of many other bacterial toxins as well as some type III secretion system effector proteins (Aktories et al., 2011; Dean, 2011). Thus, it is possible that TccC plays a role independent of its TC partners within the *Ca.* A. arthropodicus - louse fly association, again reflecting a transition from a parasitic to mutualistic lifestyle.

It is also possible that these toxin components do not provide a useful function in the symbiosis and that these candidate genes have simply not yet accumulated mutations that allow us to identify them as inactive. It is predicted that, after association with a host insect, there are many genes that are not required in the new environment that evolve under relaxed selection and are prone to accumulating nonsense mutations or insertions/deletions (Burke and Moran, 2011). However, there is anticipated to be a lag between the onset of relaxed selection and the accumulation of an inactivating mutation, such that any given genome could maintain a subset of "cryptic" pseudogenes (Burke and Moran, 2011). Given that the TcA and TcB components have already been inactivated in

Table 2.4. Homologs of TccC1 and TccC2 of *Ca.* A. arthropodicus identified using BLAST.

	Locus tag	Product	Size[a]	% ID[b]	% ID[c]
Ca. A. arthropodicus	ARA_06730	TccC1	1035		58
Ca. A. arthropodicus	ARA_04415	TccC2	860	58	
Photorhabdus luminescens	plu4167	TccC1	1043	52	58
Photorhabdus luminescens	plu4182	TccC6	965	58	57
Serratia entomophila	pADAP_57	SepC	973	58	61
Serratia proteomaculans	Spro_0382	YD repeat protein	852	57	59
Xenorhabdus bovienii	XBJ1_1574	TccC	932	59	63
Xenorhabdus nematophila	XNC1_2567	TccC	1016	47	59
Yersinia pestis	y2020	Insecticidal toxin	874	55	59
Yersinia pseudotuberculosis	YPTS_2310	YD repeat protein	994	55	59

[a] Size of product in amino acyl residues.
[b] Percent amino acid identity shared with TccC1 of *Ca.* A. arthropodicus.
[c] Percent amino acid identity shared with TccC2 of *Ca.* A. arthropodicus.

Ca. A. arthropodicus, it is possible that the remaining TcC components represent such cryptic pseudogenes and that they have no functional role in the symbiosis.

Other toxins

A CDS sharing 68% amino acid sequence identity with the *Yersinia* murine toxin (*ymt*) gene is present in *Ca.* A. arthropodicus. This murine toxin displays phospholipase D activity and is toxic to mice and rats, but has also been shown to be required for *Yersinia pestis* survival in the midgut of its flea vector (Hinnebusch et al., 2002). In *Y. pestis*, Ymt is predicted to protect the bacterium from cytotoxic products originating from the flea's digestion of its blood meal (Hinnebusch et al., 2002). Since the louse fly host of *Ca.* A. arthropodicus also feeds exclusively on vertebrate blood, it is conceivable that this gene plays a similar role.

Additionally, there are a number of candidate ORFs in *Ca.* A. arthropodicus that share homology with the repeats-in-toxin (RTX) family of proteins (Lin et al., 1999) and the RTX ABC transporter. Some of the transporter genes appear to be intact, but genes

encoding the large RTX proteins contain numerous frameshifting mutations and are broken up into many fragments. Additionally, there are small gene fragments remaining that share significant sequence identity with the *mcf* (makes caterpillars floppy) gene, which produces a toxin involved in cell lysis of insect midguts (Dowling et al., 2004) and is used by *P. luminescens* to kill insects. Identification of TC components, combined with the fragments of RTX and Mcf toxins that remain in the genome, suggest that at some point in the evolutionary past, *Ca.* A. arthropodicus played a role as an insect pathogen and that these virulence properties have been attenuated in the switch to a mutualistic relationship with the louse fly.

Type III secretion systems

Type III secretion systems (T3SSs) are commonly utilized by pathogenic bacteria to inject effector proteins directly into the cytosol of eukaryotic host cells (Hueck, 1998). These effector proteins often lead to host cytoskeletal rearrangements, allowing invasion of host cells by bacterial pathogens, or they may alter host immune response, or induce host cell apoptosis (Mota and Cornelis, 2005). Genes encoding the type III secretion apparatus, or injectisome, have been identified in divergent bacterial pathogens as well as in recently associated bacterial symbionts, and many of the structural genes are conserved between them (Cornelis, 2006). The effectors translocated via this system, however, may differ substantially between bacterial species, leading to different host effects (Dean, 2011; Mota and Cornelis, 2005). For example, there are two islands encoding T3SS genes in the tsetse symbiont, *S. glossinidius*, with synteny to T3SS islands identified in *Yersinia* and *Salmonella* spp. (Dale et al., 2001; Dale et al., 2005). In these two bacterial pathogens, the T3SS effectors suppress host immune responses and inhibit phagocytosis,

allowing the invasion of host cells, and may also lead to induction of host cell apoptosis (Cornelis, 2002; Pavlova et al., 2011). *Sodalis glossinidius* appears to use these systems in a similar manner to intitiate successful intracellular infections in a tsetse fly host, although without detrimental cytotoxic effects on host cells (Dale et al., 2001; Dale et al., 2005).

There are three islands of genes encoding components of T3SSs in the genome of *Ca.* A. arthropodicus. Like *S. glossinidius*, island 1 appears to be related to the T3SS of *Yersinia* spp., with similar gene content and organization (Fig. 2.6A), while the other two islands show synteny with the SPI-1 island found in *Salmonella* spp. (Fig. 2.6B; Dale et al., 2005; Hueck, 1998). One of the SPI-1-like islands in *Ca.* A. arthropodicus contains an inversion of the genes encoding PrgH-HrpE when compared to SP1-1, and is also interrupted between *invA* and *invB* by a phage element (Fig. 2.6B). Within each island, there are genes that have accumulated frameshifting mutations or large truncations, indicating that no single island encodes a complete secretion apparatus (Fig. 2.6A,B). However, it is possible that the individual genic components of each island function synchronously to facilitate secretion. At least one full-length copy is present of all genes predicted to be required for successful secretion, except for genes encoding the components SctV/InvA, SctU/SpaS, and SctC/InvG. The SctU and SctV proteins contain a transmembrane domain and a cytosolic domain, which are involved in formation of the export pore in the inner membrane and recognition and switching of substrates to be exported, respectively, while SctC is a secretin and is involved in formation of the channel through the outer membrane (Diepold et al., 2011). These three proteins are predicted to be necessary for formation of a functional injectisome, though in *Ca.* A.

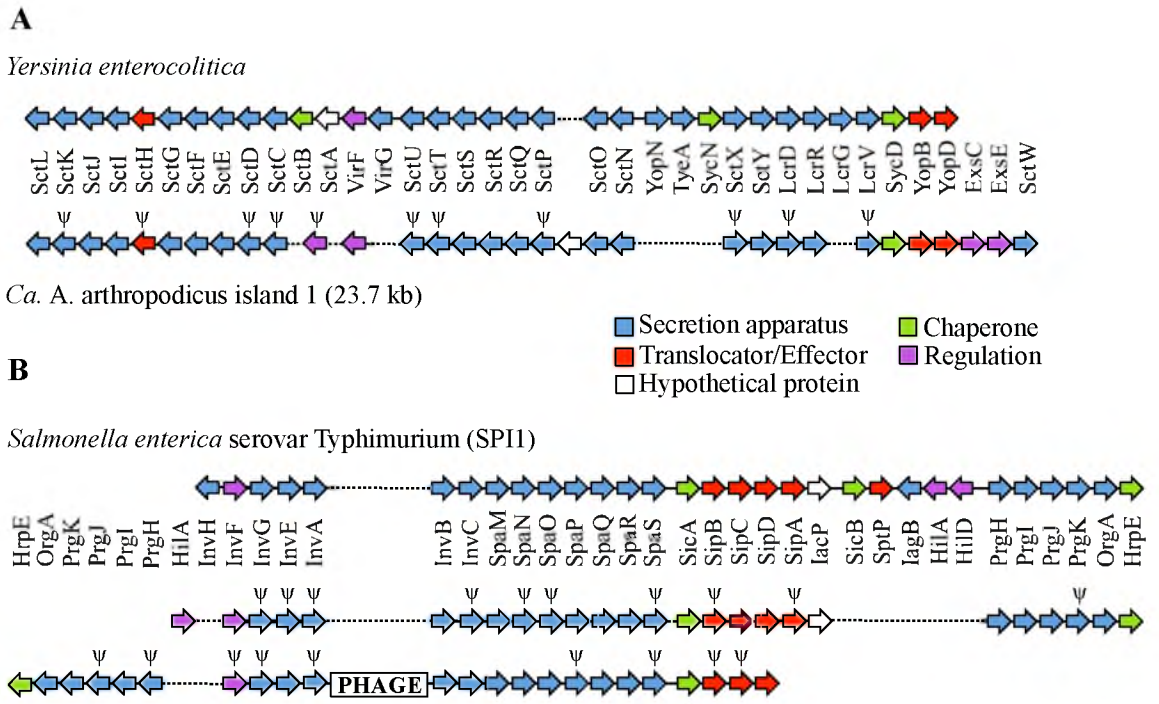


Figure 2.6. Gene content and organization of type III secretion system islands. A. The T3SS island from *Yersinia enterocolitica* compared to island 1 in *Ca.* A. arthropodicus. B. The SPI1 T3SS island from *Salmonella enterica* serovar Typhimurium compared to two islands of genes found in *Ca.* A. arthropodicus. Arrows indicate gene orientation and colors indicate the predicted role of the gene products named. Dotted lines indicate locations which contain genes in other islands. Anticipated pseudogenes are denoted by ψ.

arthropodicus, all three copies of genes encoding these products contain mutations causing frameshifts or premature stop codons.

It is possible that flagellar gene products could complement the T3SS components in *Ca.* A. arthropodicus, given that many of the structural elements of both T3SS and flagella are functionally similar (Aizawa, 2001) and the two structures are predicted to have evolved from common ancestry (Gophna et al., 2003; Saier, 2004). A recent study by Stone et al. (2010), shows that the non-motile intracellular pathogen, *Chlamydia pneumoniae*, retains three flagellar homologs, *flhA, fliF* and *fliI,* that can interact and co-

purify with certain T3SS components. The truncated T3SS proteins in *Ca.* A. arthropodicus, SctV and SctU, are structurally related to flagellar FlhA and FlhB, respectively, and an intact CDS encoding for FlhB is present in *Ca.* A. arthropodicus (Table 2.3). However, *flhA* contains a frameshifting mutation that creates a premature stop codon (Table 2.3). It is still unclear which flagellar protein SctC is structurally and functionally closest to, although the outer ring component FlgI has been suggested (Aizawa, 2001), which also contains frameshifting mutations in *Ca.* A. arthropodicus (Table 2.3). At this point it is unclear whether any of the truncated T3SS or flagellar products maintain functionality, and the requirement for an operational T3SS in *Ca.* A. arthropodicus needs to be tested using genetic approaches.

Insertion sequences and phage regions

There are 22 copies of a bacterial insertion sequence (IS) element that shares homology with IS911, which is a member of the IS3 family of IS elements. These IS elements generally consist of two adjacent, partially overlapping ORFs (Rousseau et al., 2004). BLAST homology searches show that the two consecutive ORFs in *Ca.* A. arthropodicus are closest to a 283 amino acid ORF and a 103 amino acid ORF of the IS3/IS911 family of *Shewanella denitrificans*. This IS element is repeated in seven positions in the genome in which it appears to be full-length and intact. In addition, there are 15 copies of this IS element which have similar truncations in both ORF components. These truncated elements may still retain the activity required for transposition, or may be mobilized in trans using the integrase and transposase enzymes from the intact elements.

There are a large number of repetitive phage genes present in the *Ca.* A. arthropodicus genome. Putative phage elements comprise 10% of the genome sequence and 15% of the predicted CDSs. There are phage genes sharing sequence similarity with those of the APSE bacteriophages present in *H. defensa*, a facultative symbiont of aphids (Degnan et al., 2009; Moran et al., 2005). The APSE phages in *H. defensa* carry genes encoding a variety of toxins, such as Shiga toxin, cytolethal distending toxin and YD-repeat proteins, that confer protective benefits to the aphid host in defense against parasitoid wasps (Degnan and Moran, 2008; Oliver et al., 2009). While there are multiple copies of APSE phage structural and regulatory sequences, intact toxin-encoding genes from these elements are absent in the *Ca.* A. arthropodicus genome. There is a gene fragment present that shares homology with the YD-repeat proteins, however only a small portion of the gene remains and it has likely been inactivated by deletion of a majority of the ORF. Therefore, it does not appear that *Ca.* A. arthropodicus is using these phages in the same way as *H. defensa* to protect the louse fly against parasitoid infection. However, we cannot rule out the possibility that other insecticidal proteins, such as the TccC components, provide an analogous function such that these symbionts could play a role in host defense.

Conclusion

The genome sequence of *Ca.* A. arthropodicus provides evidence of a genome in transition from a bygone parasitic lifestyle to a mutualistic associate of insects. It shares genomic traits common to secondary symbionts that have a recent host association, such as decreased inventory of functional genes and increased numbers of repetitive elements compared to free-living relatives, but still retains a larger gene inventory than insect

symbionts with an ancient origin of association, demonstrating that this transition to insect mutualism is still in progress. The presence of potential toxins and virulence factors and islands of genes encoding type III secretion system components and effectors, implies that the ancestor of this symbiont was formerly a pathogen. The switch to a mutualistic insect-associated lifestyle is predicted to have led to inactivation of toxin products and effectors that would cause direct harm to the host, and this is evidenced by the inactivated remnants of various toxins, such as the RTXs and TC components. The continued maintenance of some intact predicted virulence factors that are shared with pathogenic bacteria may indicate that mutualistic symbionts use these factors to interact with host cells using the same machinery as pathogens, but without decreasing host fitness. However, such a hypothesis would need to be clarified using genetic experiments.

The genome sequence of *Ca.* A. arthropodicus provides further insight into potential roles this symbiont might play in its wide range of insect hosts. Within the louse fly, it is possible that this symbiont serves to supplement the host blood diet with B vitamins, given the retention of genes involved in vitamin synthesis, although the prospect of other functions cannot be ruled out. Other insect symbionts are known to perform a variety of roles in hosts outside of diet supplementation, such as increasing host defenses against parasitoids by expressing toxins with insecticidal activity (Oliver et al., 2003, 2009) or increasing host thermal tolerance to temperature stresses (Montllor et al., 2002; Russell and Moran, 2006). The genome of *Ca.* A. arthropodicus contains elements related to thermal stress and insect pathogenicity that might also be used in these manners to enhance louse fly host fitness, including numerous heat shock proteins

as well as T3SS effectors and the TccC components of the insecticidal toxin complexes. Genetic experimentation and the use of aposymbiotic lines of flies will be necessary to further elucidate if *Ca.* A. arthropodicus functions in one of these capacities similar to other symbionts or if it has a unique role within louse flies.

Since *Ca.* A. arthropodicus has been isolated in culture, the genome sequence can now be used to enable genetic testing of the utility of specific genes in the symbiosis. To date, little is known about the interactions that occur between symbiont and host on a molecular level since few insect symbionts have proved amenable to axenic culture and genetic modification. This work provides a complete gene inventory for a symbiont that can be cultured and manipulated in the lab, facilitating further study of symbiont gene functions *in vivo*.

References

Aizawa, S. I. 2001. Bacterial flagella and type III secretion systems. *FEMS Microbiol Lett*, 202:157-64.

Akman, L., A. Yamashita, H. Watanabe, K. Oshima, T. Shiba, M. Hattori, and S. Aksoy. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia. *Nat Genet*, 32:402-7.

Aktories, K., A. E. Lang, C. Schwan, and H. G. Mannherz. 2011. Actin as target for modification by bacterial protein toxins. *FEBS J*, 278:4526-43.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J Mol Biol*, 215:403-10.

Belda, E., A. Moya, S. Bentley, and F. J. Silva. 2010. Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of Sodalis glossinidius, the secondary endosymbiont of tsetse flies. *BMC Genomics*, 11:449.

Blackburn, M. B., J. M. Domek, D. B. Gelman, and J. S. Hu. 2005. The broadly insecticidal Photorhabdus luminescens toxin complex a (Tca): activity against the Colorado potato beetle, Leptinotarsa decemlineata, and sweet potato whitefly,

Bemisia tabaci. *J Insect Sci*, 5:32.

Blattner, F. R., G. Plunkett, 3rd, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, et al. 1997. The complete genome sequence of Escherichia coli K-12. *Science*, 277:1453-62.

Borodovsky, M. and J. McIninch. 1993. GeneMark: Parallel gene recognition for both DNA strands. *Comput. Chem*, 17:123-133.

Boussau, B., and M. Gouy. 2006. Efficient likelihood computations with nonreversible models of evolution. *Syst Biol*, 55:756-68.

Bowen, D., T. A. Rocheleau, M. Blackburn, O. Andreev, E. Golubeva, R. Bhartia, and R. H. ffrench-Constant. 1998. Insecticidal toxins from the bacterium Photorhabdus luminescens. *Science*, 280:2129-32.

Burke, G. R., and N. A. Moran. 2011. Massive genomic decay in Serratia symbiotica, a recently evolved symbiont of aphids. *Genome Biol Evol*, 3:195-208.

Carver, T., N. Thomson, A. Bleasby, M. Berriman, and J. Parkhill. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics*, 25:119-20.

Cicmil, N., and R. H. Huang. 2008. Crystal structure of QueC from Bacillus subtilis: an enzyme involved in preQ1 biosynthesis. *Proteins*, 72:1084-8.

Cornelis, G. R. 2002. Yersinia type III secretion: send in the effectors. *J Cell Biol*, 158:401-8.

Cornelis, G. R. 2006. The type III secretion injectisome. *Nat Rev Microbiol*, 4:811-25.

Corti, A., A. F. Casini, and A. Pompella. Cellular pathways for transport and efflux of ascorbate and dehydroascorbate. *Arch Biochem Biophys*, 500:107-15.

Dale, C., M. Beeton, C. Harbison, T. Jones, and M. Pontes. 2006. Isolation, pure culture, and characterization of "Candidatus Arsenophonus arthropodicus," an intracellular secondary endosymbiont from the hippoboscid louse fly Pseudolynchia canariensis. *Appl Environ Microbiol*, 72:2997-3004.

Dale, C., T. Jones, and M. Pontes. 2005. Degenerative evolution and functional diversification of type-III secretion systems in the insect endosymbiont Sodalis glossinidius. *Mol Biol Evol*, 22:758-66.

Dale, C., and N. A. Moran. 2006. Molecular interactions between bacterial symbionts and their hosts. *Cell*, 126:453-65.

Dale, C., S. A. Young, D. T. Haydon, and S. C. Welburn. 2001. The insect endosymbiont

Sodalis glossinidius utilizes a type III secretion system for cell invasion. *Proc Natl Acad Sci U S A*, 98:1883-8.

Dale, J. M., L. Popescu, and P. D. Karp. 2010. Machine learning methods for metabolic pathway prediction. *BMC Bioinformatics*, 11:15.

Darby, A. C., J. H. Choi, T. Wilkes, M. A. Hughes, J. H. Werren, G. D. Hurst, and J. K. Colbourne. 2010. Characteristics of the genome of Arsenophonus nasoniae, son-killer bacterium of the wasp Nasonia. *Insect Mol Biol*, 19 Suppl 1:75-89.

Dean, P. 2011. Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *FEMS Microbiol Rev*, 35:1100-25.

Degnan, P. H., T. E. Leonardo, B. N. Cass, B. Hurwitz, D. Stern, R. A. Gibbs, S. Richards, and N. A. Moran. 2010. Dynamics of genome evolution in facultative symbionts of aphids. *Environ Microbiol*. 12:2060-2069.

Degnan, P. H., and N. A. Moran. 2008. Diverse phage-encoded toxins in a protective insect endosymbiont. *Appl Environ Microbiol*.

Degnan, P. H., Yu, Y., Sisneros, N., Wing, R.A., and N.A. Moran. 2009. Hamiltonella defensa, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proc Natl Acad Sci U S A*, 106:9063-9068.

Delcher, A. L., Hormon, D., Kasif, S., White, O. and S. L. Salzberg. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res*, 27: 4636-4641.

Diepold, A., U. Wiesand, and G. R. Cornelis. 2011. The assembly of the export apparatus (YscR,S,T,U,V) of the Yersinia type III secretion apparatus occurs independently of other structural components and involves the formation of an YscV oligomer. *Mol Microbiol*, 82:502-14.

Dowling, A. J., P. J. Daborn, N. R. Waterfield, P. Wang, C. H. Streuli, and R. H. ffrench-Constant. 2004. The insecticidal toxin Makes caterpillars floppy (Mcf) promotes apoptosis in mammalian cells. *Cell Microbiol*, 6:345-53.

Duchaud, E., C. Rusniok, L. Frangeul, C. Buchrieser, A. Givaudan, S. Taourit, S. Bocs, C. Boursaux-Eude, M. Chandler, J. F. Charles, et al. 2003. The genome sequence of the entomopathogenic bacterium Photorhabdus luminescens. *Nat Biotechnol*, 21:1307-13.

Duron, O., D. Bouchon, S. Boutin, L. Bellamy, L. Zhou, J. Engelstadter, and G. D. Hurst. 2008. The diversity of reproductive parasites among arthropods: Wolbachia do not walk alone. *BMC Biol*, 6:27.

Ewald, P. W. 1987. Transmission modes and evolution of the parasitism-mutualism

continuum. *Ann N Y Acad Sci.*, 503:295-306.

Ewing, B. and P. Green. 1998. Basecalling of automated sequencer traces using phred. II. Error probabilities. *Genome Res,* 8:186-194.

Ewing, B., Hillier, L., Wendl, M. and P. Green. 1998. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res,* 8:175-185.

ffrench-Constant, R., and N. Waterfield. 2005. An ABC Guide to the Bacterial Toxin Complexes. *Adv Appl Microbiol*, 58C:169-183.

Glasner, J. D., C. H. Yang, S. Reverchon, N. Hugouvieux-Cotte-Pattat, G. Condemine, J. P. Bohin, F. Van Gijsegem, S. Yang, T. Franza, D. Expert, et al. 2011. Genome sequence of the plant-pathogenic bacterium Dickeya dadantii 3937. *J Bacteriol*, 193:2076-7.

Gophna, U., E. Z. Ron, and D. Graur. 2003. Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. *Gene*, 312:151-63.

Gordon, D., Abajian, C. and P. Green. 1998. Consed: A graphical tool for sequence finishing. *Genome Res*, 8:195-202.

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*, 52:696-704.

Hinnebusch, B. J., A. E. Rudolph, P. Cherepanov, J. E. Dixon, T. G. Schwan, and A. Forsberg. 2002. Role of Yersinia murine toxin in survival of Yersinia pestis in the midgut of the flea vector. *Science*, 296:733-5.

Hueck, C. J. 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev*, 62:379-433.

Kanehisa, M., and S. Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*, 28:27-30.

Lang, A. E., G. Schmidt, A. Schlosser, T. D. Hey, I. M. Larrinua, J. J. Sheets, H. G. Mannherz, and K. Aktories. 2010a. Photorhabdus luminescens toxins ADP-ribosylate actin and RhoA to force actin clustering. *Science*, 327:1139-42.

Lang, A. E., G. Schmidt, J. J. Sheets, and K. Aktories. 2010b. Targeting of the actin cytoskeleton by insecticidal toxins from Photorhabdus luminescens. *Naunyn Schmiedebergs Arch Pharmacol*, 383:227-35.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. 2007. Clustal W and Clustal X version 2.0.

Bioinformatics, 23:2947-8.

Lin, W., K. J. Fullner, R. Clayton, J. A. Sexton, M. B. Rogers, K. E. Calia, S. B. Calderwood, C. Fraser, and J. J. Mekalanos. 1999. Identification of a Vibrio cholerae RTX toxin gene cluster that is tightly linked to the cholera toxin prophage. *Proc Natl Acad Sci U S A*, 96:1071-6.

Lowe, T. M. and S. R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucl. Acids Res*, 25: 955-964.

Lukashin A. and M. Borodovsky. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res*, 26:1107-1115.

Maezawa, K., S. Shigenobu, H. Taniguchi, T. Kubo, S. Aizawa, and M. Morioka. 2006. Hundreds of flagellar basal bodies cover the cell surface of the endosymbiotic bacterium Buchnera aphidicola sp. strain APS. *J Bacteriol*, 188:6539-43.

Marchler-Bauer, A., S. Lu, J. B. Anderson, F. Chitsaz, M. K. Derbyshire, C. DeWeese-Scott, J. H. Fong, L. Y. Geer, R. C. Geer, N. R. Gonzales, et al. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res*, 39:D225-9.

McCutcheon, J. P., and N. A. Moran. 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol*.

Montllor, C. B., Maxmen, A., and A.H. Purcell. 2002. Facultative bacterial endosymbionts benefit pea aphids Acyrthosiphon pisum under heat stress. *Ecol Entomol*, 27:189-195.

Moran, N. A. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A*, 93:2873-8.

Moran, N. A., P. H. Degnan, S. R. Santos, H. E. Dunbar, and H. Ochman. 2005. The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. *Proc Natl Acad Sci U S A*, 102:16919-26.

Moran, N. A., G. R. Plague, J. P. Sandstrom, and J. L. Wilcox. 2003. A genomic perspective on nutrient provisioning by bacterial symbionts of insects. *Proc Natl Acad Sci U S A*, 100 Suppl 2:14543-8.

Moriya, Y., M. Itoh, S. Okuda, A. C. Yoshizawa, and M. Kanehisa. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*, 35:W182-5.

Mota, L. J., and G. R. Cornelis. 2005. The bacterial injection kit: type III secretion systems. *Ann Med*, 37:234-49.

Moya, A., J. Pereto, R. Gil, and A. Latorre. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat Rev Genet*, 9:218-29.

Nogge, G. 1981. Significance of symbionts for the maintenance of an optional nutritional state for successful reproduction in hematophagous arthropods. *Parasitol*, 82:101-104.

Novakova, E., V. Hypsa, and N. A. Moran. 2009. Arsenophonus, an emerging clade of intracellular symbionts with a broad host distribution. *BMC Microbiol*, 9:143.

Oliver, K. M., P. H. Degnan, M. S. Hunter, and N. A. Moran. 2009. Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science*, 325:992-4.

Oliver, K. M., J. A. Russell, N. A. Moran, and M. S. Hunter. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc Natl Acad Sci U S A*, 100:1803-7.

Paul, K., G. Gonzalez-Bonet, A. M. Bilwes, B. R. Crane, and D. Blair. 2011. Architecture of the flagellar rotor. *EMBO J*, 30:2962-71.

Pavlova, B., J. Volf, P. Ondrackova, J. Matiasovic, H. Stepanova, M. Crhanova, D. Karasova, M. Faldyna, and I. Rychlik. 2011. SPI-1-encoded type III secretion system of Salmonella enterica is required for the suppression of porcine alveolar macrophage cytokine expression. *Vet Res*, 42:16.

Pearson, M. M., M. Sebaihia, C. Churcher, M. A. Quail, A. S. Seshasayee, N. M. Luscombe, Z. Abdellah, C. Arrosmith, B. Atkin, T. Chillingworth, et al. 2008. Complete genome sequence of uropathogenic Proteus mirabilis, a master of both adherence and motility. *J Bacteriol*, 190:4027-37.

Pichon, S., D. Bouchon, R. Cordaux, L. Chen, R. A. Garrett, and P. Greve. 2009. Conservation of the Type IV secretion system throughout Wolbachia evolution. *Biochem Biophys Res Commun*, 385:557-62.

Rances, E., D. Voronin, V. Tran-Van, and P. Mavingui. 2008. Genetic and functional characterization of the type IV secretion system in Wolbachia. *J Bacteriol*, 190:5020-30.

Rio, R. V., R. E. Symula, J. Wang, C. Lohs, Y. N. Wu, A. K. Snyder, R. D. Bjornson, K. Oshima, B. S. Biehl, N. T. Perna, M. Hattori, and S. Aksoy. 2012. Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: glossinidae) obligate symbiont wigglesworthia. *MBio*, 3.

Robb, F. T., D. L. Maeder, J. R. Brown, J. DiRuggiero, M. D. Stump, R. K. Yeh, R. B.

Weiss, and D. M. Dunn. 2001. Genomic sequence of hyperthermophile, Pyrococcus furiosus: implications for physiology and enzymology. *Methods Enzymol*, 330:134-57.

Rousseau, P., E. Gueguen, G. Duval-Valentin, and M. Chandler. 2004. The helix-turn-helix motif of bacterial insertion sequence IS911 transposase is required for DNA binding. *Nucleic Acids Res*, 32:1335-44.

Russell, J. A., and N. A. Moran. 2006. Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. *Proc Biol Sci*, 273:603-10.

Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell. 2000. Artemis: sequence visualization and annotation. *Bioinformatics*, 16:944-5.

Saier, M. H., Jr. 2004. Evolution of bacterial type III protein secretion systems. *Trends Microbiol*, 12:113-5.

Shigenobu, S., H. Watanabe, M. Hattori, Y. Sakaki, and H. Ishikawa. 2000. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. *Nature*, 407:81-6.

Sorfova, P., A. Skerikova, and V. Hypsa. 2008. An effect of 16S rRNA intercistronic variability on coevolutionary analysis in symbiotic bacteria: molecular phylogeny of Arsenophonus triatominarum. *Syst Appl Microbiol*, 31:88-100.

Steinert, M., U. Hentschel, and J. Hacker. 2000. Symbiosis and pathogenesis: evolution of the microbe-host interaction. *Naturwissenschaften*, 87:1-11.

Stone, C. B., D. C. Bulir, J. D. Gilchrist, R. K. Toor, and J. B. Mahony. 2010. Interactions between flagellar and type III secretion proteins in Chlamydia pneumoniae. *BMC Microbiol*, 10:18.

Tatusov, R. L., E. V. Koonin, and D. J. Lipman. 1997. A genomic perspective on protein families. *Science*, 278:631-7.

Toft, C., and M. A. Fares. 2008. The evolution of the flagellar assembly pathway in endosymbiotic bacterial genomes. *Mol Biol Evol*, 25:2069-76.

Toh, H., B. L. Weiss, S. A. Perkin, A. Yamashita, K. Oshima, M. Hattori, and S. Aksoy. 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of Sodalis glossinidius in the tsetse host. *Genome Res*, 16:149-56.

Waterfield, N. R., T. Ciche, and D. Clarke. 2009. Photorhabdus and a host of hosts. *Annu Rev Microbiol*, 63:557-74.

Weeks, A. R., M. Turelli, W. R. Harcombe, K. T. Reynolds, and A. A. Hoffmann. 2007. From parasite to mutualist: rapid evolution of Wolbachia in natural populations of Drosophila. *PLoS Biol*, 5:e114.

Watson, W. T., T. D. Minogue, D. L. Val, S. B. von Bodman, and M. E. Churchill. 2002. Structural basis and specificity of acyl-homoserine lactone signal production in bacterial quorum sensing. *Mol Cell*, 9:685-94.

Wilkes, T. E., Duron, O., Darby, A. C., Hypsa, V., Novakova, E., and G. D. D. Hurst. 2011. The genus Arsenophonus. In E. Zchori-Fein, and K. Bourtzis (ed.), *Manipulative tenants: Bacteria associated with arthropods*. CRC Press, Boca Raton, FL.

CHAPTER 3

COMPARATIVE ANALYSIS OF THE LOUSE FLY SYMBIONT, CANDIDATUS ARSENOPHONUS ARTHROPODICUS, AND THE PARASITOID WASP SYMBIONT, ARSENOPHONUS NASONIAE

Abstract

The *Arsenophonus* clade of bacteria consists of symbionts associated with a wide range of insect hosts having different diets and lifestyles. These symbiotic bacteria maintain a wide array of interactions with these hosts ranging from mutualistic to parasitic and vertically to horizontally transmitted. We have compared the genome sequences of two members of the *Arsenophonus* clade: the mutualistic louse fly symbiont, *Candidatus* Arsenophonus arthropodicus and the reproductive parasite of parasitoid wasps, *Arsenophonus nasoniae*. These two symbionts are closely related and have genomes that share a high level of synteny, although they reside in different insect hosts with distinct host effects. Analysis of gene inventory and pseudogene content suggests that both of these symbionts have become associated with insects recently, although *Ca.* A. arthropodicus has undergone more degenerative evolution than *A. nasoniae*. Compared to free-living relatives, both *Arsenophonus* symbionts have lost regulatory elements, likely due to their lifestyles within the static insect host environment, and have accumulated pseudogenes and repetitive elements, similar to other

recently associated insect symbionts. Given the diversity in the range of associations displayed by *Arsenophonus* symbionts, this group presents an interesting system with which to study the molecular mechanisms of insect interactions.

Introduction

Many microbial genome sequencing projects have been completed over the past decade, providing a wealth of information on a wide variety of bacterial species, including many bacterial symbionts of insects. Comparative genomics provides a means to use this data to understand their evolutionary histories and the role of bacterial symbionts in their hosts. Here we compare the genome sequences of two members of the *Arsenophonus* genus, which contains species that have now been identified as associates in a wide range of distantly related insect hosts, including plant-feeding insects such as aphids and psyllids, and vertebrate blood-feeders including ticks, triatomines and flies (Duron et al., 2008; Novakova et al., 2009). *Arsenophonus* species have been classified as both obligate, bacteriome-associated symbionts, as well as recently acquired facultative symbionts, and different species may undergo vertical or horizontal transmission (Wilkes et al, 2011). Owing to the diversity of association types involving *Arsenophonus* species and differences observed in transmission strategies in their insect hosts, this genus presents an interesting group of bacteria with which to investigate the mechanistic basis of insect symbiosis. There are two species of *Arsenophonus* that have been isolated in axenic culture and represent the first members of this group to have undergone genome sequencing: *Arsenophonus nasoniae*, a symbiont of a parasitoid wasp (Darby et al., 2010; Gherna et al., 1991; Werren et al., 1986), and *Candidatus* Arsenophonus arthropodicus, a symbiont of the pigeon louse fly (Dale et al., 2006).

These two closely related symbionts inhabit insects that have different lifestyles and the bacteria seem to have distinct effects on their hosts, making them useful for comparing genes involved in insect association.

Arsenophonus nasoniae is a bacterial symbiont associated with the Hymenopteran parasitoid wasp, *Nasonia vitripennis*. Parasitoid wasps, such as *N. vitripennis,* parasitize cyclorrhaphous Dipteran flies, ovipositing their own offspring into host fly pupae (Danneels et al., 2010). Prior to oviposition, the wasps inject venom into the fly pupa, which leads to immune suppression and eventual cell apoptosis and fly death, allowing the wasp larvae to develop and subsist on the fly pupa (Danneels et al., 2010). Only a proportion of *N. vitripennis* wasps maintain populations of the symbiont *A. nasoniae*, but wasps that are infected inject the bacteria along with the eggs into the fly pupal host. *Arsenophonus nasoniae* then colonizes and replicates within the parasitized fly pupa (Werren et al., 1986; Wilkes et al., 2010). When the developing wasp larvae feed on the pupa, they ingest the symbiotic bacteria in the process, which then pass through the gut to colonize the wasp (Darby et al., 2010). While this mechanism is how *A. nasoniae* is vertically transmitted from mother to offspring, it also provides a means of horizontal transfer to offspring of uninfected parents, if the same fly pupa is parasitized by multiple wasps (Duron et al., 2010; Huger et al., 1985).

Nasonia vitripennis wasps that have acquired an infection of *A. nasoniae* exhibit a sex ratio bias in their offspring toward females, with males being killed as embryos (Gherna et al., 1991; Werren et al., 1986). *Nasonia vitripennis* is a haplodiploid insect, and this male killing occurs as a result of the inability of the maternally-derived centrosomes to form in unfertilized (male) eggs (Ferree et al., 2008). This type of male-

killing is predicted to be caused by production of a small molecule or effector that can cross the wasp egg membrane and disrupt the formation of the maternal centrosome (Ferree et al., 2008; Wilkes et al., 2010). Thus, *A. nasoniae* effects host reproductive distortions similar to those observed in insects with *Wolbachia* infections (Bandi et al., 2001).

Candidatus Arsenophonus arthropodicus is a symbiont of the Dipteran hippoboscid louse fly, *Pseudolynchia canariensis*. The louse fly is an obligate blood-feeding parasite of rock pigeons (*Columbia livia*) that spends the majority of its life on its bird host, with females leaving only to deposit offspring in the nest material or to transfer to a new host (Baequart, 1953; Marshall, 1981). *Candidatus* Arsenophonus arthropodicus resides in a variety of tissues in the louse fly, including the hemolymph, gut, fat body and reproductive tissues (Dale et al., 2006). This symbiont is found in both male and female louse flies and does not appear to have any reproductive effects on its host, with an observed sex ratio of 50:50 (Dale et al., 2006).

The louse fly is related to the tsetse fly, and both reproduce by means of adenotrophic viviparity, in which a single egg is fertilized at a time and all larval stages occur within the uterus of the mother, with the larva obtaining nutrition through milk gland secretions (Attardo et al., 2008). The tsetse fly maintains two different symbionts: a bacteriome-associated primary symbiont, *Wigglesworthia glossinidia* (Aksoy, 1995), and a second symbiont found in multiple fly tissues, *Sodalis glossinidius* (Dale and Maudlin, 1999). Like tsetse, the louse fly also harbors an obligate primary symbiont, in addition to *Ca.* A. arthropodicus, that is closely related to *W. glossinidia* (Dale et al., 2006). The louse fly primary symbiont likely plays a nutritive role in its host, similar to

that of *W. glossinidia*, which plays a role in dietary supplementation and blood meal digestion (Akman et al., 2002; Nogge, 1981; Pais et al., 2008). Both of the louse fly symbionts are transmitted to offspring vertically through the milk gland secretions that larvae feed on while developing *in utero*, in a manner similar to symbiont transmission in tsetse flies (Attardo et al., 2008). In addition, *Ca.* A. arthropodicus symbionts have no known life stage component outside of the louse fly host.

Here we compare the genome sequence of *Ca.* A. arthropodicus with the draft genome sequence of *A. nasoniae* (Darby et al., 2010; Wilkes et al., 2010) with the aim of understanding how these two closely related bacteria interact with hosts that have distinct lifestyles. We note that both symbionts have relatively recent origins of association with insect hosts, although the genome sequence of *Ca.* A. arthropodicus appears to be more degenerate, with a smaller genome size and gene inventory than *A. nasoniae*. We identify factors that may be involved in insect associations that are shared by both species as well as factors that differ between the two associations and may account for the abilities of these bacteria to infect unique insect hosts with different outcomes.

Materials and Methods

Genome alignments

The 143 individual *A. nasoniae* draft genome scaffold sequences deposited in GenBank (Darby et al., 2010) were aligned with the *Ca.* A. arthropodicus genome sequence using progressiveMauve (Darling et al., 2004; Darling et al., 2010). A single GenBank file was then produced for *A. nasoniae* by concatenating the individual scaffold GenBank files in their aligned order using the union tool in Emboss (Rice et al., 2000).

The two *Arsenophonus* genomes were then compared using Artemis v13.2.0 (Rutherford et al., 2000) and DNAPlotter (Carver et al., 2009).

Genome analyses

Conceptually translated amino acid fasta files for both *Arsenophonus* genomes were compared using BLAST (Altschul et al., 1990) and BLASTCLUST to identify genes that shared high levels of sequence identity. Regions of genome synteny were identified using Crossmatch, which is a component of the Phred/Phrap/Consed assembly package (Ewing and Green, 1998; Ewing et al., 1998; Gordon et al., 1998). Plots of genome synteny were then generated using Circos (Krzywinski et al., 2009). Metabolic capabilities were compared using the KEGG (Kyoto Encyclopedia of Genes and Genomes; Kanehisa and Goto, 2000) automatic annotation server (KAAS; Moriya et al., 2007).

Pseudogene analysis

The *Ca.* A. arthropodicus and *A. nasoniae* genome scaffolds were each independently aligned to the genome sequence of a closely related outgroup, *Proteus mirabilis* (Pearson et al., 2008) using progressiveMauve (Darling et al., 2010) to determine genome synteny. Orthologs sharing at least 50% identity over 70% of their length were exported and the sizes of the intact *P. mirabilis* orthologs were used to compare pseudogene sizes to the sizes of intact coding sequences.

A simple Monte Carlo approach was developed to simulate the evolution of pseudogenes in *Ca.* A. arthropodicus. The program simulates the accumulation of random mutations in all of the orthologs of intact genes and pseudogenes that are shared

by *Ca.* A. arthropodicus and its closest free-living relative, *Proteus mirabilis*. Mutations accumulate in accordance with ORF size in a randomly selected class of neutral genes over a number of cycles. At preset cycle intervals, the simulation records (i) the difference in size between intact and disrupted sequences, (ii) the number of neutral genes that have accumulated one or more disrupting mutations, and (iii) the density of disrupting mutations.

Results and Discussion

General features

The *A. nasoniae* genome is currently a draft genome sequence that comprises 143 scaffolds (Darby et al., 2010), while the *Ca.* A. arthropodicus genome sequence is composed of a single scaffold (Table 3.1). *Arsenophonus nasoniae* has a genome size that is larger than *Ca.* A. arthropodicus, and has a larger gene inventory, implying that *Ca.* A. arthropodicus has undergone more extensive degenerative evolution over the course of its association with the louse fly (Table 3.1). To identify genes shared between the two symbionts, BLASTCLUST analysis of the two genomes was used to cluster together coding sequences (CDSs) that shared amino acid sequence identity with cutoffs at 75% and 95% identity. Out of the 2203 identified CDSs predicted to encode functional gene products in *Ca.* A. arthropodicus, 66% of them paired with a CDS in *A. nasoniae* with which they shared at least 75% amino acid sequence identity or higher and 26% were found to share over 95% identity (Fig. 3.1). The remaining CDSs shared amino acid sequence identity below 75% or were unique to one of the *Arsenophonus* genomes.

Excluding phage-associated CDSs and repetitive elements, *Ca.* A. arthropodicus contains just 37 unique chromosomal CDSs that are intact and do not have apparent

Table 3.1. Features of the *Ca.* A. arthropodicus and *A. nasoniae* genome sequences.

	Ca. A. arthropodicus	*A. nasoniae*
Size (Chromosome + Plasmids)	2,971,954 bp	3,567,128 bp
GC content (%)	38.1	37.4
# of scaffolds (# of contigs)	1 (3)	144 (665)
Predicted CDSs	2,231	3,203
Pseudogenes	395	135
rRNA operons	7	8-10[a]
tRNAs	64	52
Plasmids	3	2 or more[a]
% Prophage sequence	10	10

[a] Estimates based on fragmented assembly.

homologs in the *A. nasoniae* genome sequence (Table 3.2). Of these 37 CDSs, nine of them would be located between scaffolds or within sequence gaps in their predicted location in the *A. nasoniae* sequence (Table 3.2). Ten of the CDSs are located in a block in *Ca.* A. arthropodicus (ARA_06845-ARA_06530) and share homology with nine consecutive CDSs in *Proteus mirabilis*. There is one interspersing CDS in this block that is predicted to encode a cysteine desulfurase enzyme that is involved in thiamine biosynthesis, and shares homology with that of *Sphaerobacter thermophilus*, a distantly related bacterium, suggesting an independent acquisition of this enzyme. There is also a block comprising seven consecutive *Ca.* A. arthropodicus CDSs (ARA_10960-ARA_10995) that is absent in the *A. nasoniae* sequence, encoding products predicted to be involved in lipopolysaccharide (LPS) biosynthesis (Table 3.2). The loss of multiple adjacent CDSs that are present in both *Ca.* A. arthropodicus and closely related bacteria, such as *P. mirabilis* and *Photorhabdus luminescens*, suggests that *A. nasoniae* has undergone several large deletions. However, for the most part, *Ca.* A. arthropodicus appears to retain a subset of the *A. nasoniae* gene set, with few unique protein coding genes that are not shared by both symbionts.

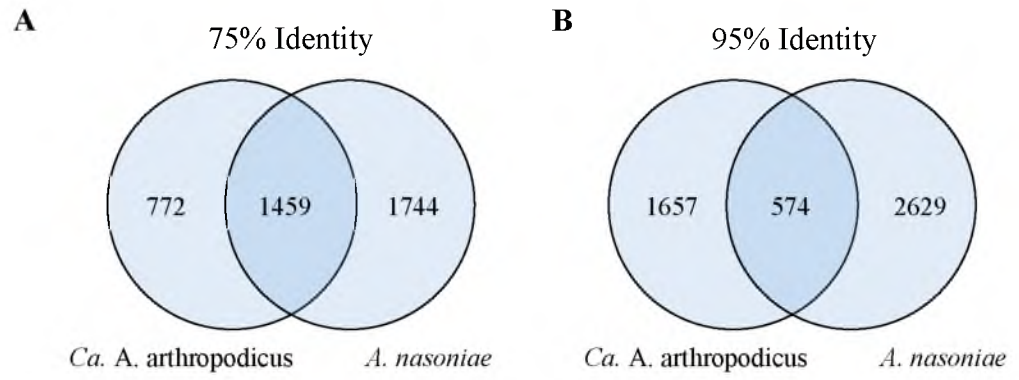


Figure 3.1. BLASTCLUST analysis of intact *Ca.* A. arthropodicus CDSs sharing amino acid sequence identity with CDSs in *A. nasoniae*. A.Venn diagram showing the number of CDSs with protein sequence identity of 75% or higher. B. CDSs with greater than 95% identity.

Genome alignment and synteny

To determine the level of synteny shared between the two genomes, *A. nasoniae* scaffolds that were at least 5 kb in size were aligned to the *Ca.* A. arthropodicus sequence. These alignments were then compared using Crossmatch, with a minscore of 100 and remaining default parameters (Gordon et al., 1998), which determined the presence of CDSs that were located in the same order and orientation within each genome sequence. There were 491 alignments obtained between the two genome sequences with an average length of 5,002 bp, average nucleotide identity of 88.5% and a total alignment length of 2,456,051 bp. This output was used to generate a plot of synteny (Fig. 3.2) using Circos (Krzywinski et al., 2009). The Circos alignment shows that, within the scaffolds of *A. nasoniae*, there is a high level of synteny with fragments of the *Ca.* A. arthropodicus genome, indicating that many gene orthologs are retained in the same order and orientation in each organism.

Table 3.2. Unique CDSs in the *Ca.* A. arthropodicus genome sequence.

Locus Tag	Product	BLAST homology - Locus tag
ARA_02740	Hypothetical protein	*Photorhabdus asymbiotica* - PAU02164
ARA_04590	Hypothetical protein	*Hamiltonella defensa* - Hdef_1253
ARA_04830	CDP-diacylglycerol pyrophosphatase	*Burkholderia* sp. - bgla_1g12250
ARA_06360	Hypothetical protein	*Proteus mirabilis* - PMI1624
ARA_06845	Citrate lyase beta	*Proteus mirabilis* - PMI0231
ARA_06490	Siderophore biosynthesis	*Proteus mirabilis* - PMI0232
ARA_06495	TonB siderophore receptor	*Proteus mirabilis* - PMI0233
ARA_06500	Diaminopimelate decarboxylase	*Proteus mirabilis* - PMI0234
ARA_06505	Pyridoxal-phosphate dep enzyme	*Proteus mirabilis* - PMI0235
ARA_06510	octopine/opine dehydrogenase	*Proteus mirabilis* - PMI0236
ARA_06515	MFS-family transporter	*Proteus mirabilis* - PMI0237
ARA_06520	Iron ABC transporter	*Proteus mirabilis* - PMI0238
ARA_06525	Cysteine desulfurase	*Sphaerobacter thermophilus* - Sthe3332
ARA_06530	Hypothetical protein	*Proteus mirabilis* - PMI_0239
ARA_07095[a]	Fumarate reductase Fe-S subunit	*Proteus mirabilis* - PMI3587
ARA_07110[a]	Fumarate reductase subunit D	*Proteus mirabilis* - PMI3585
ARA_07725	Murine toxin	*Yersinia pestis* - y1069
ARA_07820	Hypothetical protein	*Pseudomonas syringae* - Psyr_2271
ARA_09180	Xylose isomerase	*Burkholderia* sp. - BC1001_0668
ARA_09190	Multidrug efflux transporter	*Burkholderia* sp. - BC1001_0669
ARA_09195	Glycosyltransferase family 2 protein	*Burkholderia* sp. - BC1001_0670
ARA_09795	Phosphoesterase	*Proteus mirabilis* - PMI3123
ARA_10960	LPS biosynthesis protein	*Photorhabdus luminescens*- plu4812
ARA_10965	Aminotransferase	*Photorhabdus luminescens*- plu4813
ARA_10970	LPS biosynthesis protein	*Photorhabdus luminescens*- plu4814
ARA_10975	Glycosyltransferase family protein	*Shewanella baltica* - Sbal195_3023
ARA_10980	Hypothetical protein	*Escherichia coli* - ECED1_2379
ARA_10990	Glycosyltransferase family protein	*Haemophilus parasuis* - HPS_03029
ARA_10995	Glycosyltransferase family protein	*Camnocytophaga ochracea* - Coch0695
ARA_11150	Prevent-host-death protein	*Pectobacterium carotovorum*
ARA_11245[a]	Phosphate ABC transport- pstS	*Proteus mirabilis* - PMI_2893
ARA_11250[a]	Phosphate ABC transport- pstC	*Proteus mirabilis* - PMI_2894
ARA_12095[a]	MFS-family transporter	*Photorhabdus luminescens*- plu0476
ARA_12100[a]	Phosphoglycolate phosphatase	*Photorhabdus luminescens*- plu0475
ARA_12105[a]	Alcohol dehydrogenase	*Photorhabdus luminescens*- plu0474
ARA_12135[a]	DNA-directed DNA polymerase	*Proteus mirabilis* - PMI2485
ARA_15258[a]	PTS system- Hpr	*Photorhabdus luminescens*- plu1394

[a] Predicted location in *A. nasoniae* genome contains sequence gaps or scaffold breaks.

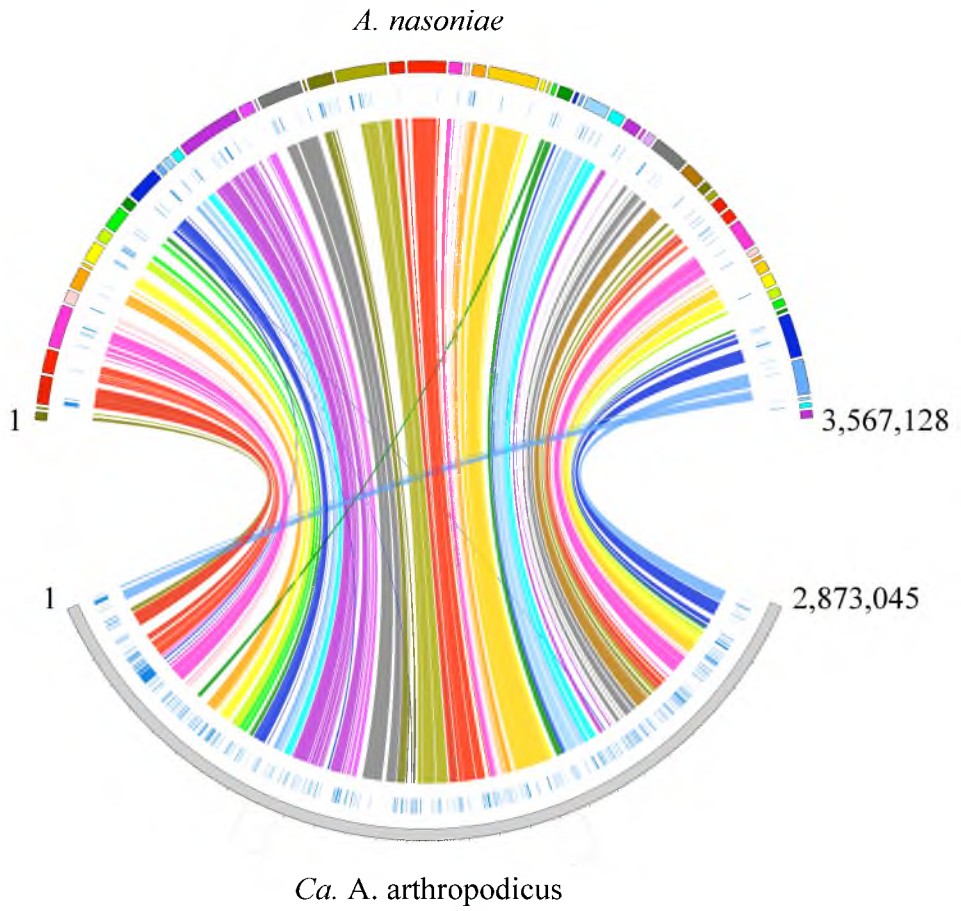


Figure 3.2. Plot of genome synteny between *A. nasoniae* and *Ca.* A. arthropodicus. Individual scaffolds of the *A. nasoniae* genome are indicated by the outermost colored boxes along the top and aligned to the concatenated *Ca.* A. arthropodicus genome sequence represented by the gray bar along the bottom. Colored bands joining the two represent individual areas of synteny along the lengths the genomes. The track of blue lines along each genome represent the numbers and positions of pseudogenes.

Due to the fragmented nature of the *A. nasoniae* genome sequence, we cannot predict the order and orientation of scaffolds and in Fig. 3.2, these scaffolds are ordered and oriented according to the *Ca.* A. arthropodicus assembly. Chromosomal rearrangements may have occurred in either or both species that distort synteny at the level of the whole genome sequence. To examine the likelihood that the *A. nasoniae* scaffold alignment also shares this high level of overall synteny with the *Ca.* A.

arthropodicus genome, the GC skew of the *A. nasoniae* genome was plotted using DNAPlotter (Carver et al., 2009) based on the scaffold alignment obtained from progressiveMauve (Darling et al., 2010). Both *Arsenophonus* genomes show a similar polarized GC skew that is typical of bacterial genomes (Fig. 3.3), with polarity switching around the replication origin and terminus due to asymmetry in the number of G residues versus C residues present in the leading and the lagging strands (Arakawa and Tomita, 2007). Therefore, the *A. nasoniae* genome scaffolds may align in an orientation similar to that of the *Ca.* A. arthropodicus genome sequence, and these two species appear to maintain a high level of overall genome synteny, suggesting a recent divergence.

Pseudogene analysis

Bacterial symbionts are known to undergo extensive genome degeneration following restriction in an insect host (Dale and Moran, 2006). Analysis of symbiont genome sequences shows that insect symbionts with an ancient origin of association have lost a large percentage of genes, leading to a streamlining of the genome to contain only genes necessary to maintain the symbiosis (Dale and Moran, 2006). It has been predicted that after becoming associated with an insect host, there are many genes that may evolve under relaxed selection because they are not required in the static insect environment and are free to accumulate inactivating mutations without detrimental effect (Burke and Moran, 2011). However, there is predicted to be a lag between the time of insect association, when selection becomes relaxed on a subset of non-required genes, and the time required for all of these genes to accumulate noticeable inactivating mutations (Burke and Moran, 2011). This means that in the early stages of insect association there can exist so-called "cryptic" pseudogenes that are evolving under relaxed selection and

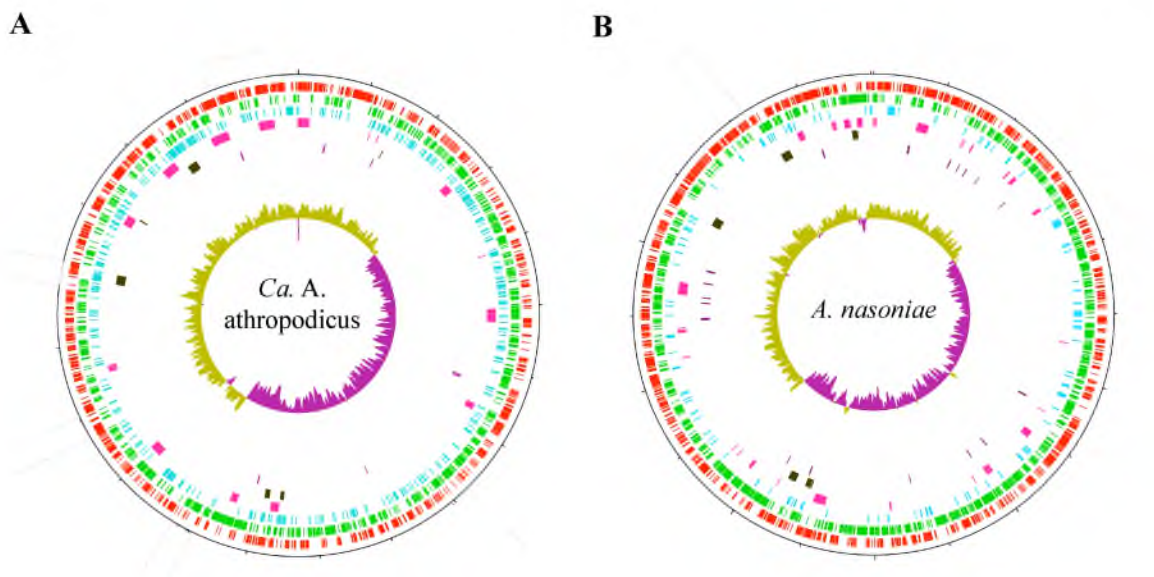


Figure 3.3. GC skew and positions of CDSs and pseudogenes in *A. nasoniae* and *Ca.* A. arthropodicus chromosomes. From outer track: genes encoded on the leading strand (red), genes on lagging strand (green), pseudogenes (light blue), phage genes (pink), type III secretion system genes (brown), two-component and quorum sensing regulatory elements (purple). Innermost circle depicts GC skew.

are therefore functionally redundant but have not yet acquired any inactivating mutations. Given a constant random mutation rate and a Poisson distribution of gene sizes, it can be predicted that within a subset of genes evolving under relaxed selection, larger genes are expected to accumulate inactivating mutations more rapidly than smaller genes. As time passes, the likelihood of any given gene accumulating a nonsense or frameshifting mutation increases until all of the genes under relaxed selection have accumulated one or more inactivating mutations. However, in the interim, larger genes are expected to preferentially accumulate such mutations. Therefore, if we compare the average sizes of genes that have acquired inactivating mutations to those that remain intact, we expect to see a significant increase in the average sizes of pseudogenes in comparison to intact genes early in the symbiotic association. This size difference is then expected to decrease

over time as more genes acquire mutations, and this can provide an estimate of how far a symbiont is on the trajectory of genome degeneration that occurs following obligate host association.

To analyze the pseudogene content of *Ca.* A. arthropodicus and *A. nasoniae*, we first aligned the two *Arsenophonus* genome sequences with that of their closest fully sequenced free-living relative, *Proteus mirabilis*, using progressiveMauve (Darling et al., 2010). Orthologs shared between each of the *Arsenophonus* species and *P. mirabilis* were then exported for analysis. To determine the average sizes of genes that remain intact versus those that contain one or more inactivating mutations in each *Arsenophonus* species, the sizes of their orthologs present in *P. mirabilis* were used, since genes that have been inactivated may have become truncated in the degenerative process and may not accurately reflect their ancestral state. Out of a subset of 1468 orthologs shared between *Ca.* A. arthropodicus and *P. mirabilis*, 165 of them contained inactivating mutations in *Ca.* A. arthropodicus. The average size of the *Ca.* A. arthropodicus pseudogene orthologs is 1267 bp, while the average size of the remaining intact genes is 960 bp (Fig. 3.4). Thus, the size difference between the pseudogene orthologs and the intact orthologs is 307 bases. For *A. nasoniae*, the pseudogene orthologs had an average size of 1496 bp and intact genes had an average size of 981 bp. The difference in size between orthologs of pseudogenes vs. intact genes of *A. nasoniae* was larger than that of *Ca.* A. arthropodicus, with *A. nasoniae* pseudogenes, on average, being 515 bases longer than intact genes. These results suggest that both *Arsenophonus* symbionts have recent origins of association with insects, although *Ca.* A. arthropodicus seems to have been host-restricted longer than *A. nasoniae* and has undergone more substantial genome

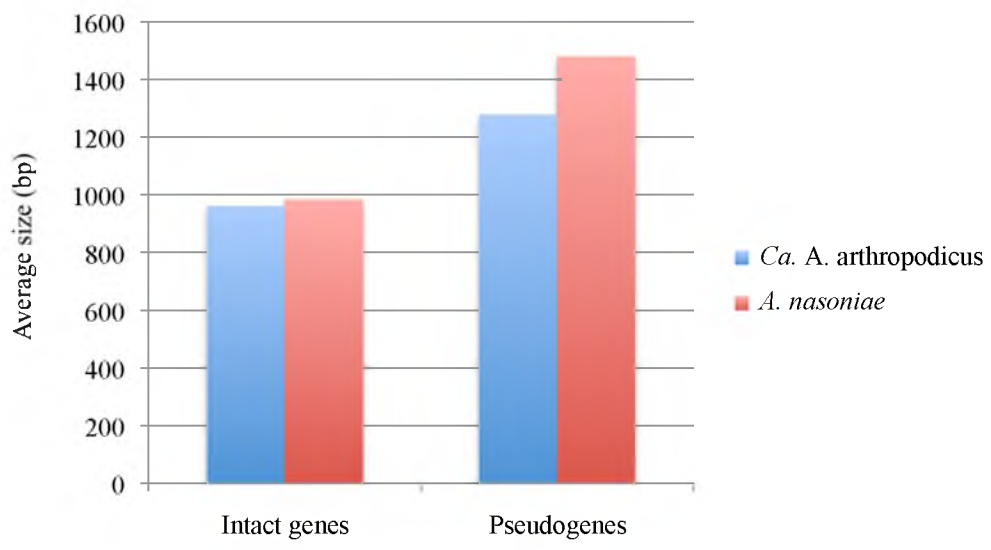


Figure 3.4. Average sizes of pseudogenes and intact orthologs for *A. nasoniae* and *Ca.* A. arthropodicus. Average sizes were calculated based on the sizes of orthologs in the close free-living relative, *P. mirabilis.* For *Ca.* A. arthropodicus intact genes, n=1303; pseudogenes, n=165. For *A. nasoniae* intact genes, n=1408; pseudogenes, n=33.

degeneration, leading to an increase in the number of inactivated genes present and a decrease in the size differential between pseudogenes and the intact genes.

We attempted to estimate the number of "cryptic" pseudogenes present *Ca.* A. arthropodicus, using a Monte Carlo simulation that is parameterized by the subset of *P. mirabilis* genes that share orthologs in *Ca.* A. arthropodicus (1468 orthologs). In the simulation, genes were permitted to randomly accumulate mutations at a constant rate for a specified number of mutational cycles and the numbers of genes that acquired mutations vs. those that remained intact, as well as their average sizes, were computed at intervals. We ran multiple iterations of this simulation with different parameters for the number of genes that would be immediately under relaxed selection after insect association, in order to estimate the number of "cryptic" pseudogenes in *Ca.* A. arthropodicus. Using the simulation with 1468 *P. mirabilis* orthologs, a predicted

population of 500 genes under initial relaxed selection yields an average size difference of 303 bp when 167 genes have acquired inactivating mutations (Fig. 3.5), which closely matches the empirical numbers observed for *Ca.* A. arthropodicus (165 inactivated orthologs, 307 bp size difference). In this simulation, only a third of the 500 potential genes have acquired a mutation, therefore, we can predict that there are an additional 333 "cryptic" pseudogenes within the subset of orthologs that are currently evolving under relaxed selection and are eligible to acquire inactivating mutations. Since we only used a subset of *Ca.* A. arthropodicus genes in this simulation, we can scale this proportion up to obtain an estimate for the number of genes under relaxed selection in the complete genome. If we assume, based on the simulation results, that one third of the pool of available genes have acquired mutations, and *Ca.* A. arthropodicus has 395 ORFs annotated as pseudogenes in the complete genome, we can predict that there are approximately 1185 total genes under relaxed selection, with 790 "cryptic" pseudognes that could be inactivated and lost in the near future. However, this estimate is based on only a subset of the predicted genes in the genome of *Ca.* A. arthropodicus. Identifying a larger set of orthologs by manual analysis of all the remaining CDSs in the *Ca.* A. arthropodicus genome might lead to an improved estimate of the number of "cryptic" pseudogenes present that are under relaxed selection. It should also be noted that our model relies on the assumption that the mandate for genes under relaxed selection is prescribed at the outset of the symbiotic association. However, it seems likely that changes in the interactions between the host and symbiont over the course of their association could result in the relaxation of selection on genes that were formerly adaptive.

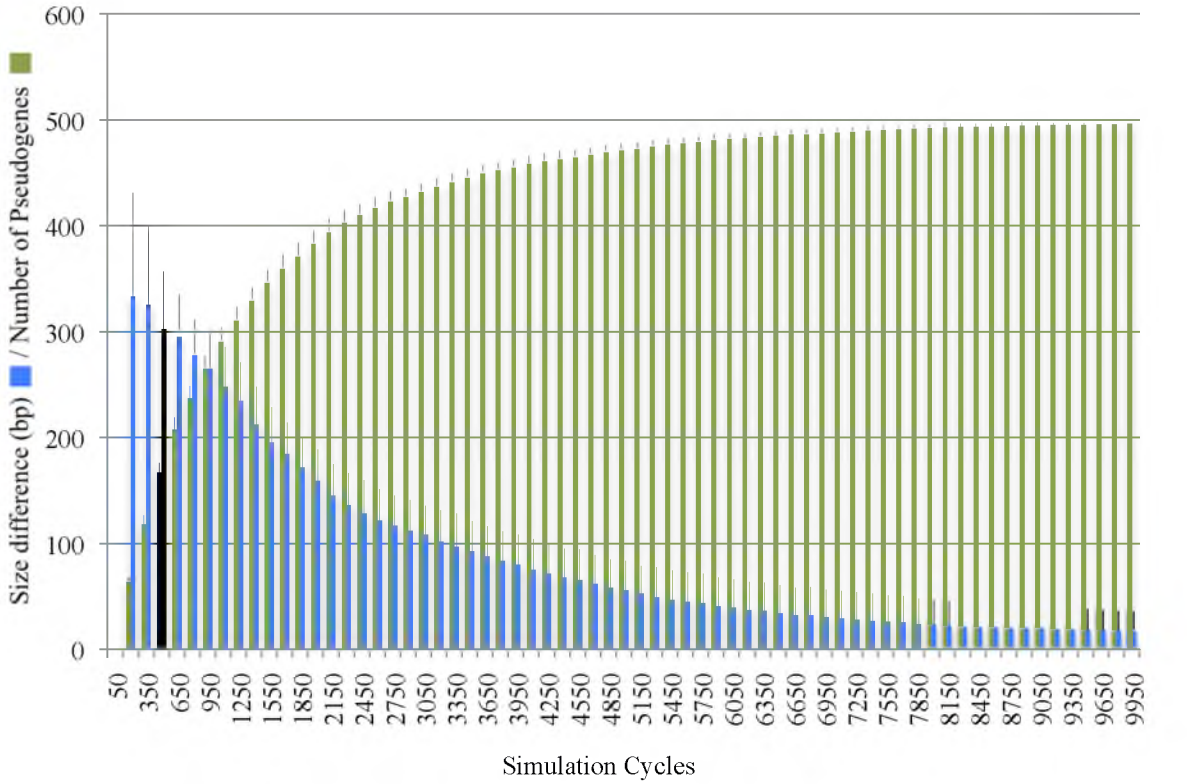


Figure 3.5. Monte Carlo simulation of gene inactivation in a subset of *Ca.* A. arthropodicus orthologs. Results of simulation for *Ca.* A. arthropodicus, based on a gene set of 1468 *P. mirabilis* orthologs, with 500 orthologs estimated to be evolving under relaxed selection immediately after association with an insect host. Blue bars represent the difference in base pairs between the average length of pseudogenes and the average length of intact genes and green bars indicate the number of genes that have acquired an inactivating mutation. The set of black bars indicate the position at which the *Ca.* A. arthropodicus genome falls. The X axis represents time, with increasing numbers of mutational cycles (generations). Each bar represents the average value based on 10 simulation runs, with error bars representing one standard deviation.

Plasmids

The plasmid content of *A. nasoniae* is not fully assembled; however, it appears to maintain two or more plasmids, one of which is small (c. 8 kb) and has been visualized by gel electropohoresis (Darby et al., 2010). This small plasmid may be similar to pARA3, a 9,893 bp plasmid identified in *Ca.* A. arthropodicus which maintains only a few intact CDSs encoding a LuxR-like regulatory element, a cytosine permease and a

colicin immunity protein, all of which have homologs in small scaffolds within the *A. nasoniae* genome sequence. Thus, both symbionts may maintain similar versions of this plasmid.

Arsenophonus nasoniae is predicted to carry one or more additional larger plasmids with genes encoding products involved in conjugative transfer that may be similar to plasmid pARA1 in *Ca.* A. arthropodicus. The 55,855 bp plasmid, pARA1, contains a set of genes sharing homology with the type IV secretion system (T4SS) *tra* genes involved in conjugative transfer (Murata et al., 2002; Pearson et al., 2008), however, a number of the genes required for this process have accumulated inactivating mutations in *Ca.* A. arthropodicus. *A. nasoniae*, however, appears to maintain a more intact set of the genes required for conjugative transfer in multiple genome locations (Darby et al., 2010). Thus, conjugal transfer may play a role in the lifestyle of *A. nasoniae*.

A third plasmid present in *Ca.* A. arthropodicus, pARA2, contains genes with sequence similarity to components of the VirB-like T4SS found in *Agrobacterium tumefaciens* (Fronzes et al., 2009). Homologs of these *vir*-like T4SS genes were not identified in the *A. nasoniae* genome sequence, indicating that this plasmid may have been acquired independently by *Ca.* A. arthropodicus. Similar T4SSs have been identified in both Gram-positive and Gram-negative bacteria, and play a variety of roles, including conjugative transfer of DNA, DNA uptake from the environment and delivery of effector molecules into host cells (Fronzes et al., 2009). Therefore, this T4SS may play a role analagous to the inactivated conjugative transfer genes on pARA1 or this

system may function in a different manner, such as in effector secretion, as in *A. tumefaciens* (Cascales and Christie, 2003; Fronzes et al., 2009).

Two-component regulatory systems

Two-component regulatory systems are common in bacteria and consist of a histidine kinase protein that senses an environmental variable and a cognate response regulator protein that then regulates transcription of target genes (Stock et al., 2000). These regulatory systems are often involved in responding to nutrient availability or environmental stresses, and in many bacteria may play important roles in virulence or symbiosis (Stock et al., 2000). Both *Arsenophonus* symbionts retain a smaller inventory of regulatory elements, such as two-component systems, compared to free-living relatives (Table 3.3). This is predicted to be due to their static lifestyle within insects, while free-living bacteria must respond to environmental changes more frequently and may be more dependent upon functional regulatory systems that enable rapid changes in gene expression as conditions warrant. Analysis using the KEGG Automatic Annotation Pathyway (KAAS; Moriya et al., 2007) of both *Arsenophonus* genomes indicates that there are 11 complete two-component systems present in *A. nasoniae*, five of which are shared with *Ca.* A. arthropodicus (Table 3.3). These shared systems include CpxA/CpxR, which is involved in protein folding, biofilm formation and cell envelope stress response (Dorel et al., 2006), KdpD/KdpE, involved in potassium sensing and the osmotic stress response (Vlisidou et al., 2010), ArcA/ArcB, involved in the oxidative stress response and anaerobic respiration (Georgellis et al., 2001; Loui et al., 2009), RcsC/RcsD, involved in colanic acid capsule synthesis and biofilm formation, (Garcia-

Table 3.3. Presence of two-component regulatory systems in bacterial species based on KEGG pathway analysis.

	Escherichia coli K-12	*Dickeya dedantii*	*Proteus mirabilis*	*Photorhabdus luminescens*	*Arsenophonus nasoniae*	*Ca.* Arsenophonus arthropodicus	*Sodalis glossinidius*	*Wigglesworthia glossinidia*	*Buchnera aphidicola*
PhoB/PhoR	+	+	+	+	+				
PhoP/PhoQ	+	+	+	+	PhoP	PhoQ	+		
OmpR/EnvZ	+	+	+	+	+	EnvZ	+		
RstA/RstB	+	+							
CpxR/CpxA	+	+	+	+	+	+	+	CpxR	
CreB/CreC	+								
BaeR/BaeS	+	+	+	+	+	BaeR			
BasR/BasS	+	+	BasR				+		
CusR/CusS	+	+					+		
QseB/QseC	+		QseC						
KdpE/KdpD	+	+	+	+	+	+	KdpE		
TorR/TorS	+								
ArcA/ArcB	+	ArcA	+	+	+	+	+		
TctD/TctE		+		+					
NarL/NarX	+	NarX	+	NarX					
NarP/NarQ	+		NarP	NarP	NarP	NarP			
UhpA/UhpB	+	+	+	+					
RcsD/RcsC	+	+	+	+	+	+	+		
UvrY/BarA	+	+	UvrY	+	UvrY	UvrY	+		
EvgA/EvgS	+			+	+				
FimZ	+								
GlnG/GlnL	+	+	+	+	+	GlnG	+		
HydG/HydH	+								
AtoC/AtoS	+								
PgtA/PgtB			+						
CitB/CitA	+	+							
DcuR/DcuS	+	+							
QseF/QseE	+	+	+	+	+	+	QseF		
TtrR/TtrS			+						
CheA/CheY	+	+	+	+	+				
# TCSs [a]	**27**	**18**	**15**	**15**	**11**	**5**	**9**	**0**	**0**

[a] Number of complete two-component systems (TCSs) encoded. (+) indicates a complete TCS; partial TCSs are indicated by the name of the component present.

Calderon et al., 2007), and QseF/QseE, involved in regulating virulence and metabolism (Reading et al., 2009).

In addition to the shared two-component systems, *A. nasoniae* maintains six additional two-component systems which are absent or contain inactivating mutations in one or both components in *Ca.* A. arthropodicus (Table 3.3). These include systems involved in responding to phosphate and iron availability (PhoB/PhoR; Chakraborty et al., 2011), nitrogen availability (GlnL/GlnG; Reitzer, 2003), osmotic stress (EnvZ/OmpR; Yuan et al., 2011), multidrug resistance and phenotypic variation (EvgA/EvgS; Derzelle et al., 2004a; Eguchi et al., 2004), chemotaxis (CheA/CheY; Baker et al., 2006) and envelope stress response and multidrug resistance (BaeS/BaeR; Baranova and Nikaido, 2002; Raffa and Raivio, 2002). The loss of additional two-component regulatory elements in *Ca.* A. athropodicus provides further evidence that this symbiont has undergone greater genome degeneration than *A. nasoniae*. This may be due to the fact that *Ca.* A. athropodicus has been isolated for a longer period of time within a single insect host, under more constant conditions, while *A. nasoniae* is known to switch between at least two different insect environments, which may provide different nutrient availabilities, stresses and gene expression requirements.

Interestingly, neither *A. nasoniae* nor *Ca.* A. arthropodicus encodes a complete PhoP/PhoQ two-component system (Table 3.3), which senses Mg^{2+} and Ca^{2+} availability and regulates virulence genes in many bacterial pathogens (Groisman, 2001). This regulatory system is common to insect pathogens such as *Photorhabdus luminescens*, in which it is required for virulence in insect hosts (Derzelle et al., 2004b). The PhoP-PhoQ system is also known to play an important role in allowing the recently-associated tsetse

symbiont, *S. glossinidius,* to infect insect hosts by regulating genes involved in resistance to antimicrobial peptides (Pontes et al., 2011). PhoP mutants of *S. glossinidius* were unable to initiate successful infections in tsetse fly and louse fly hosts, indicating the necessity of this response regulator gene in the symbiotic association (Pontes et al., 2011). However, the PhoQ sensor kinase of *S. glossinidius* seems to lack an environmental sensing capability, indicating that it may be undergoing degeneration as an adaptation to its symbiotic lifestyle within the insect host (Pontes et al., 2011). In *A. nasoniae,* a complete *phoP* gene is present but the *phoQ* sensor kinase gene contains a frameshifting mutation, while *Ca.* A. arthropodicus maintains an intact *phoQ* gene but only a small fragment of *phoP* remains. These *Arsenophonus* symbionts are therefore expected to express genes allowing them to survive insect host immune responses and exposure to antimicrobial peptides independently of the PhoP/PhoQ two-component system.

Quorum sensing

Both *Ca.* A. arthropodicus and *A. nasoniae* maintain genes involved in quorum sensing, which is a type of two-component system found in wide variety of bacteria, by which genes are regulated based on bacterial population density (Waters and Bassler, 2005). The tsetse symbiont, *S. glossinidius,* encodes an acyl-homoserine lactone (AHL)-based quorum sensing system, composed of an AHL synthase (the LuxI homolog, SogI), which is involved in synthesis of a single AHL molecule, and a cognate response regulator (the LuxR homolog, SogR1), that is predicted to play a role in expression of genes involved in oxidative stress response (Pontes et al., 2008). Both *Arsenophonus* symbionts have a LuxI/LuxR-like quorum sensing system found in similar orientations

within their genomes. Additionally, *A. nasoniae* encodes a second pair of *luxI*/*luxR* genes that is not present in *Ca.* A. arthropodicus. This additional LuxI homolog shares 49% amino acid sequence identity with the former, suggesting independent ancestry. Since similar LuxI AHL synthases can produce different AHL molecules (Watson et al., 2002), it is conceivable that these two systems produce (and respond to) different signals, and may perform separate functions.

Adjacent to the second LuxI/LuxR pair in *A. nasoniae* is another LuxR-like quorum sensing transcriptional regulator with an unannotated ORF adjacent to it that shares sequence homology with GCN5 acetyltransferase proteins of *Pantoea ananatis*. These acetyltrasferases are also located adjacent to genes encoding LuxR regulatory proteins in *P. ananatis*. This pair of genes is also found in *Ca.* A. arthropodicus, and an acetyltransferase is also located adjacent to *sogR2*, an additional *luxR* homolog identified in *S. glossinidius* (Pontes et al., 2008). LuxI-like AHL synthases belong to the GCN5 family of acetyltransferases and share high levels of structural similarity with these acetyltransferases (Watson et al., 2002; Williams, 2007). Therefore, it is possible that this represents an additional undescribed quorum sensing system that is amenable to experimental investigation. Expression of the various *luxI* genes in *E. coli* and thin-layer chromotography overlay assays with a biosensor bacterial strain, such as *A. tumefaciens* (Shaw et al., 1997), could be used to determine the number and types of AHL molecules produced by these symbionts.

Acyl-homoserine lactone based quorum sensing systems have not yet been identified in ancient insect symbionts, such as *B. aphidicola* (Shigenobu et al., 2000) and *W. glossinidia* (Akman et al., 2002), further supporting the notion of a recent origin for

the *Arsenophonus* symbionts. Additionally, it is notable that *Ca.* A. arthropodicus maintains fewer AHL-based quorum sensing systems than *A. nasoniae*. Even among other recent symbionts that maintain quorum sensing systems, such as *S. glossinidius* (Toh et al., 2006), only a single functional system appears to be present, so it is interesting that *A. nasoniae* encodes two or three putatively functional AHL-based quorum sensing systems. Given its lifestyle in multiple host insects, regulating genes based on its population numbers may play an important role in allowing it to coordinate gene expression levels in different hosts. It will be interesting to examine how these multiple regulatory systems interact to control gene expression and their requirements within the symbiotic associations.

In addition to the AHL-based quorum sensing systems, both *Arsenophonus* symbionts carry a homolog of the *luxS* gene, which is involved in synthesis of the autoinducer-2 (AI-2) molecule (Xavier and Bassler, 2003). The AI-2 system is prevalent throughout both Gram-positive and Gram-negative bacteria and plays a role in cell metabolism, biofilm formation, and may be involved in interspecies communication, given the widespread distribution of the *luxS* gene (Vendeville et al., 2005; Xavier and Bassler, 2003). The *luxS* gene has not been identified in *S. glossinidius,* but is present in *P. luminescens*, where it regulates a wide variety of genes involved in metabolism, motility, oxidative stress response and virulence during the early stages of insect infection (Krin et al., 2006). Thus, this system may also play an important role in gene regulation and insect association in the *Arsenophonus* symbionts.

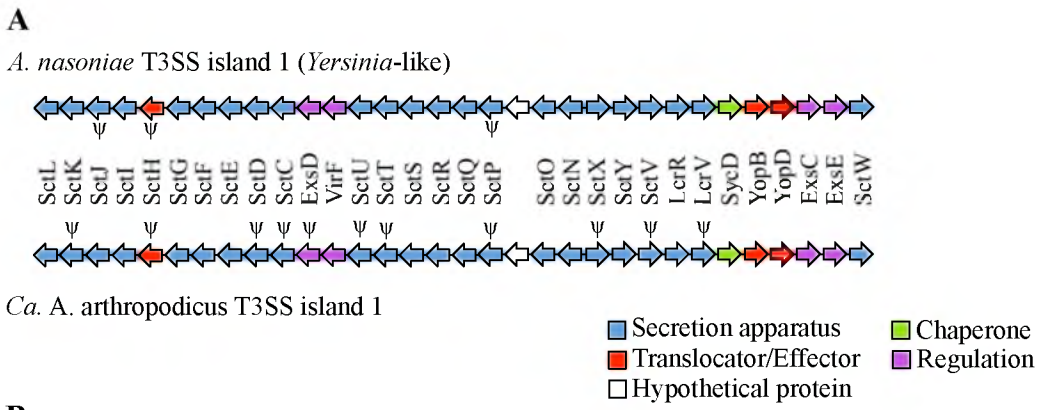
Flagella/chemotaxis

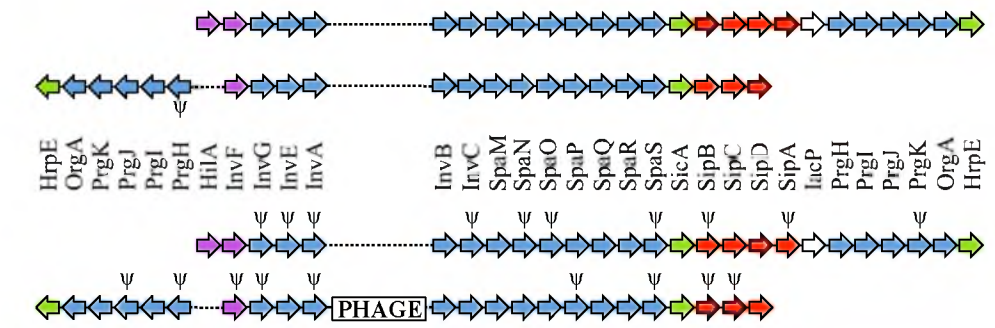
Both *Arsenophonus* symbionts carry genes encoding products involved in the regulation and synthesis of flagella. The flagellar gene inventory for *A. nasoniae* is larger and more intact (Darby et al., 2010) than that of *Ca.* A. arthropodicus, although sequencing gaps prevent a complete assessment of the status of all flagellar genes in *A. nasoniae*. The flagellar gene set in *Ca.* A. arthropodicus contains many genes that have been truncated or have accumulated frameshifting mutations. It is unclear at this point whether the flagellar genes still maintain some function in *Ca.* A. arthropodicus, either in motility or protein secretion or whether flagella do not play a vital role in its association with its louse fly host.

In addition to the flagella-encoding genes, *A. nasoniae* maintains an operon containing eight genes involved in chemotaxis (Darby et al., 2010), which are absent in *Ca.* A. arthropodicus. Chemotaxis is a means of directing motility toward or away from a chemical stimulus, such as up a nutrient concentration gradient (Bren and Eisenbach, 2000). Perhaps this ability to sense environmental gradients and respond appropriately using flagellar motility is necessary for *A. nasoniae* to successfully infect its multiple insect hosts, but is not required in the association between *Ca.* A. arthropodicus and its louse fly host.

Type III secretion systems

Both *Arsenophonus* symbiont genomes contain multiple islands of sequences encoding components of type III secretion systems (T3SS). There are three distinct islands of T3SS genes in *Ca.* A. arthropodicus which are also present in *A. nasoniae* (Fig. 3.6A,B). Similar to T3SS genes in *S. glossinidius* (Dale et al., 2005), one of the islands is





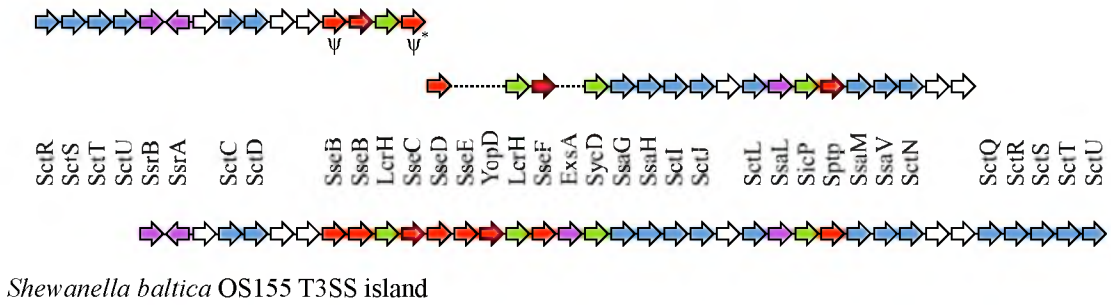


Figure 3.6. Type III secretion system islands in *A. nasoniae* and *Ca.* A. arthropodicus. Arrows indicate gene orientation and colors indicate the predicted role of the gene products named. Dotted lines indicate locations that contain genes in other islands. Anticipated pseudogenes are denoted by ψ. ψ* indicates a predicted pseudogene that contains both a mutation and a sequence gap that requires validation.

similar in gene content and organization to the *Yersinia* T3SS (Fig. 3.6A) and two of the islands are similar to the Inv/Spa SPI-1 island in *Salmonella* (Fig. 3.6B; Hueck, 1998). Within these islands, *A. nasoniae* maintains a much greater proportion of intact genes than *Ca.* A. arthropodicus, in which many of the genes in each island have acquired mutations (Fig. 3.6A,B). In addition to these shared islands, *A. nasoniae* also maintains two unique fragments of T3SS-encoding islands. These two fragments share homology with the T3SS island of *Shewanella baltica*, but have undergone some rearrangements and are located in separate locations in the genome (Fig. 3.6C). These *Shewanella*-like fragments contain additional copies of genes encoding structural proteins as well as additional effector ptoteins not present in *Ca.* A. arthropodicus. The presence of multiple predominantly intact T3SS islands in *A. nasoniae* suggests this recently associated symbiont may use these secretion systems to interact with host cells in a manner similar to that of other symbionts, such as *S. glossinidius*, or bacterial pathogens, while the loss of many of these elements in *Ca.* A. arthropodicus indicates that these systems may be undergoing degeneration and are not as integral for insect interaction in this symbiont.

Toxins: ymt, insecticidal, aip

The *ymt* gene encodes the *Yersinia* murine toxin, which has been shown to play a vital role in allowing *Yersinia pestis* to colonize the midgut of its host flea vector. Without Ymt, *Y. pestis* is affected by cytotoxic byproducts from the flea's digestion of its blood meal and cannot sustain an infection in the midgut (Hinnebusch et al., 2002). A homolog of the *ymt* gene is present in the genome of *Ca.* A. arthropodicus but absent in the genome of *A. nasoniae*, despite the fact that its putative location is fully represented in the fractured assembly (Fig. 3.7). The hosts of *A. nasoniae* do not feed on blood, so

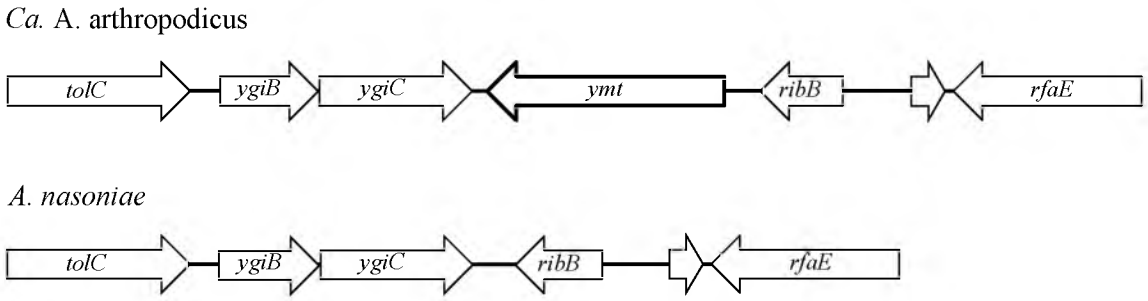


Figure 3.7. Schematic illustrating the *ymt* gene in *Ca.* A. arthropodicus and its corresponding location in *A. nasoniae*. Gene presence, size and orientation are indicated by arrows.

this gene might not play an important role in these interactions if it has a similar function to that described in *Y. pestis*. Thus, it may have been under relaxed selection and lost from the *A. nasoniae* genome. However, the presence of *ymt* in the gene inventory in *Ca.* A. arthropodicus may have played an important role in allowing *Arsenophonus* to infect blood feeding insects, such as the louse fly. It will be interesting to examine whether the presence or absence of this gene is associated with the insect host diet for other insects harboring *Arsenophonus*. For example, the dog tick, *Dermatocentor variabilis,* and the triatomine bug, *Triatoma infestans*, are two other insects that harbor *Arsenophonus* species and feed on blood (Grindle et al. 2003; Hypsa and Dale, 1997). If this gene is important in protecting symbionts from byproducts of host blood meal digestion, it should also be retained in these symbionts if they are associated with the midgut and expected to come into contact with the toxic byproducts of blood digestion.

The presence of genes sharing homology with insecticidal toxin complex (TC) genes in both *Arsenophonus* species implies that these symbiotic relationships arose from an insect pathogenic ancestor. The TCs were first identified in *P. luminescens* (Bowen et

al., 1998), and have since been found in a variety of insect pathogens (Waterfield et al., 2001), and they require three different components for complete toxicity that is active against a wide range of insects (ffrench-Constant and Waterfield, 2005). Both *A. nasoniae* and *Ca.* A. arthropodicus maintain a single copy of genes encoding two of the three components necessary for an active toxin complex, the TcdA and TcdB components, and encode two copies of the third component, TccC (ffrench-Constant and Waterfield, 2005). Both symbionts have accumulated frameshifting mutations in the TcdA and TcdB component-encoding genes of the toxin complexes. In *A. nasoniae*, both copies of the *tccC* genes have also acquired frameshifting mutations, while *Ca.* A. arthropodicus maintains two full length ORFs encoding these components. The TccC products have been shown to have ADP-ribosylating activity, like many other toxins and T3SS effectors (Aktories et al., 2011), and may function in *P. luminescens* by causing actin polymerization in target host cells, leading to host cell lysis and insect death (Lang et al., 2010a; 2010b). Given that the ORFs encoding the *tccC* genes are much larger than the average gene length in these *Arsenophonus* species (the two intact copies in *Ca.* A. arthropodicus have sizes of 3105 bp and 2580 bp, while the average pseudogene size is 1267 bp) and that both copies in *A. nasoniae* have accumulated inactivating mutations, it is tempting to speculate that these gene products are providing some useful function within the *Ca.* A. arthropodicus - louse fly association since they have not yet been inactivated, although this hypothesis will need to be experimentally investigated. It is also possible that they currently represent "cryptic" pseudogenes and, as such, provide no useful function in the symbiosis.

Arsenophonus nasoniae also has four copies of a gene that shares homology with genes producing apoptosis inducing proteins (AIP), specifically, AIP56 from *Photobacterium damselae* subsp. *piscicida* (*Phdp*; do Vale et al., 2005). *Phdp* is a pathogen of many species of fish, and depends on expression of AIP56 to evade phagocytosis by inducing apoptosis in fish host macrophages and neutrophils (do Vale et al., 2005; Silva et al., 2010). The AIP proteins encoded by *A. nasoniae* are of a similar length to those found in *Phdp* and orthologs are not present in *Ca.* A. arthropodicus or *P. mirabilis*. Three of the four gene copies are associated with phage elements in the genome, while the fourth copy is adjacent to the *Shewanella*-like T3SS, indicating that these genes may have been acquired independently in *A. nasoniae* through lateral gene transfer. At this point it is unclear what role these AIP toxins might play in the association of *A. nasoniae* with its insect hosts.

Polyketide synthesis

There are a few ORFs in *A. nasoniae* predicted to encode products that are involved in polyketide synthesis and efflux (Wilkes et al., 2010). If this is a functional system, it may present a means by which *A. nasoniae* can produce small molecules that could play a role in the male killing of *N. vitripennis* wasps (Wilkes et al., 2010). These ORFs are not present in *Ca.* A. arthropodicus, indicating that they may perform unique functions in the associations *A. nasoniae* has with its insect hosts and their functionality should be further investigated.

Conclusion

The comparative analyses presented here suggest that members of the *Arsenophonus* species have only recently become associated with insect hosts. This is evidenced by the fact that although they maintain genomes that are slightly smaller than their free-living relatives, their genome sizes and gene inventories are still large relative to symbionts that share ancient associations with insects (McCutcheon and Moran, 2012). Furthermore, analysis of the pseudogene content of these genomes also suggests that they have only recently become host-associated. Ancient insect symbionts usually possess very small genomes with relatively few pseudogenes, since the majority of neutral genes have already been inactivated and deleted (McCutcheon and Moran, 2012). Both *Arsenophonus* species maintain a larger number of pseudogenes than commonly observed in ancient symbionts, although they retain a reduced number of obvious pseudogenes relative to the tsetse symbiont, *S. glossinidius*, which has a pseudogene content that comprises an estimated 50% of its coding capacity (Belda et al., 2010; Toh et al., 2006). The *S. glossinidius* genome is predicted to contain 1501 pseudogenes (Belda et al., 2010), compared to the 395 we identified in *Ca.* A. arthropodicus and 135 in *A. nasoniae* (Darby et al., 2010), suggesting that *S. glossinidius* has been subject to a more extensive period of genome degeneration arising from obligate host-association. This leads to the prediction that both *Ca.* A. arthropodicus and *A. nasoniae* have a more recent association with their insect hosts.

In both of the *Arsenophonus* symbionts, it is notable that the average size of genes with inactivating mutations (pseudogenes) is significantly larger than the average size of all genes in the genomes, leading to the prediction that there are many genes evolving

neutrally that have not yet acquired any disrupting substitutions. This again favors the notion that the *Arsenophonus* symbionts have only recently become associated with their insect hosts. Over time, this size difference between inactivated pseudogenes and intact genes is predicted to decrease as more genes accumulate disrupting mutations. Comparison of the two *Arsenophonus* genomes indicates that *Ca.* A. arthropodicus has progressed further along a degenerative trajectory and has lost a greater number of genes from its ancestral inventory, and contains a larger number of pseudogenes. It is clear in the analyses presented here, *Ca.* A. arthropodicus has a reduced gene inventory in comparison to *A. nasoniae* in many of the categories described, including two-component and quorum sensing regulatory systems, type III secretion systems, and flagellar motility and chemotaxis. Additionally, the increase in the average size of pseudogenes relative to all genes is significantly greater in *A. nasoniae*, further supporting the notion that the *A. nasoniae* symbiosis has a more recent origin than the association between *Ca.* A. arthropodicus and the louse fly.

The presence of remnants or intact ORFs encoding insecticidal toxin complexes and other toxic products and the maintenance of T3SS structural proteins and effectors in both *Arsenophonus* symbionts indicates that they may have evolved from an ancestor that was an insect pathogen. Analysis of gene content and homology supports a close relationship between the *Arsenophonus* symbionts and the insect pathogen, *P. luminescens*, suggesting that mutualistic relationships between *Arsenophonus* spp. and their insect hosts may have arisen from pathogenic interactions. Toxic effects resulting from these virulence gene products seem to have been attenuated in the associations involving *Arsenophonus* symbionts, although the toxins and secretion systems that

remain may be participating in benign or mutualistic interactions with their insect hosts rather than antagonistic interactions that have detrimental effects on host fitness. Both *Ca.* A. arthropodicus and *A. nasoniae* retain different inventories of these virulence factors, indicating that they may play different roles in their respective insect associations. Notably, *Ca.* A. arthropodicus is only known to interact with a single insect host in a mutualistic manner, while *A. nasoniae* must successfully infect and establish within multiple insect hosts, with different outcomes.

Due to the range of insect associations and host effects they display, the *Arsenophonus* group of bacteria present an interesting study system (Wilkes et al., 2011). The biology of the two *Arsenophonus* species presented here varies greatly, with each having a different lifestyle, transmission strategy and host effects. *Arsenophonus nasoniae* has male-killing effects on its wasp host and can easily undergo horizontal transfer, while *Ca.* A. arthropodicus seems to maintain a mutualistic relationship with the louse fly and is vertically transmitted directly to offspring in a manner similar to the tsetse fly symbiont, *S. glossinidius*. It may be that *Ca.* A. arthropodicus plays a role in the louse fly that is more similar to that which *S. glossinidius* plays in tsetse flies than that of the reproductive parasite, *A. nasoniae,* in wasps. It has been predicted that *S. glossinidius* may play a role in supplementing the blood diet of tsetse flies, consistent with the retention of genes involved in vitamin synthesis (Toh et al., 2006). Very close relatives of *S. glossinidius* are primary symbionts of weevils, and they are also known to produce vitamins to enhance their insect host's fitness (Heddi et al., 1999). Additionally, different hippoboscid fly species have been described that harbor symbionts that are allied to both the *Arsenophonus* and *Sodalis* clades (Novakova and Hypsa, 2007;

Trowbridge et al., 2006), further suggesting that these two fly symbionts may be functionally interchangeable.

Now that we have a comparison between two *Arsenophonus* symbionts and have identified a number of genes of interest, it will be quite useful to test the necessity of specific genes on each of the associations. This may help to determine the functions of symbionts within their insect hosts and which genes are important in initiating and maintaining an intimate relationship. It will also be interesting to conduct reciprocal symbiont swaps to determine the ability of each symbiont to establish an infection in the other's host insects and determine if any host reproductive effects are observed. A better understanding of how each symbiont behaves in different host insects might shed light on the roles of unique vs. shared genes in these *Arsenophonus* species and uncover which genes are responsible for observed host effects, such as male-killing.

Acknowledgments

This manuscript will be submitted for publication with Diane M. Dunn, Robert B. Weiss, Alexander T. Aoyagi, Kelly F. Oakeson, Abhishek Chari, Alvin J. Clark and Colin Dale as coauthors.

References

Akman, L., A. Yamashita, H. Watanabe, K. Oshima, T. Shiba, M. Hattori, and S. Aksoy. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia. *Nat Genet*, 32:402-7.

Aksoy, S. 1995. Wigglesworthia gen. nov. and Wigglesworthia glossinidia sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. *Int J Syst Bacteriol*, 45:848-51.

Aktories, K., A. E. Lang, C. Schwan, and H. G. Mannherz. 2011. Actin as target for modification by bacterial protein toxins. *FEBS J*, 278:4526-43.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J Mol Biol*, 215:403-10.

Arakawa, K., and M. Tomita. 2007. The GC skew index: a measure of genomic compositional asymmetry and the degree of replicational selection. *Evol Bioinform Online*, 3:159-68.

Attardo, G. M., N. Guz, P. Strickler-Dinglasan, and S. Aksoy. 2006. Molecular aspects of viviparous reproductive biology of the tsetse fly (Glossina morsitans morsitans): regulation of yolk and milk gland protein synthesis. *J Insect Physiol*, 52:1128-36.

Attardo, G. M., C. Lohs, A. Heddi, U. H. Alam, S. Yildirim, and S. Aksoy. 2008. Analysis of milk gland structure and function in Glossina morsitans: milk protein production, symbiont populations and fecundity. *J Insect Physiol*, 54:1236-42.

Baker, M. D., P. M. Wolanin, and J. B. Stock. 2006. Signal transduction in bacterial chemotaxis. *Bioessays*, 28:9-22.

Bandi, C., A. M. Dunn, G. D. Hurst, and T. Rigaud. 2001. Inherited microorganisms, sex-specific virulence and reproductive parasitism. *Trends Parasitol*, 17:88-94.

Belda, E., A. Moya, S. Bentley, and F. J. Silva. 2010. Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of Sodalis glossinidius, the secondary endosymbiont of tsetse flies. *BMC Genomics*, 11:449.

Bequaert, J. The Hippoboscidae or louse-flies (Diptera) of mammals and birds. Part I. Structure, physiology and natural history. 1953. *Entomol Am* 32:1-209.

Bowen, D., T. A. Rocheleau, M. Blackburn, O. Andreev, E. Golubeva, R. Bhartia, and R. H. ffrench-Constant. 1998. Insecticidal toxins from the bacterium Photorhabdus luminescens. *Science*, 280:2129-32.

Bren, A., and M. Eisenbach. 2000. How signals are heard during bacterial chemotaxis: protein-protein interactions in sensory signal propagation. *J Bacteriol*, 182:6865-73.

Burke, G. R., and N. A. Moran. 2011. Massive genomic decay in Serratia symbiotica, a recently evolved symbiont of aphids. *Genome Biol Evol*, 3:195-208.

Carver, T., N. Thomson, A. Bleasby, M. Berriman, and J. Parkhill. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics*, 25:119-20.

Cascales, E., and P. J. Christie. 2003. The versatile bacterial type IV secretion systems. *Nat Rev Microbiol*, 1:137-49.

Chakraborty, S., J. Sivaraman, K. Y. Leung, and Y. K. Mok. 2011. Two-component PhoB-PhoR regulatory system and ferric uptake regulator sense phosphate and iron to control virulence genes in type III and VI secretion systems of Edwardsiella tarda. *J Biol Chem*, 286:39417-30.

Dale, C., M. Beeton, C. Harbison, T. Jones, and M. Pontes. 2006. Isolation, pure culture, and characterization of "Candidatus Arsenophonus arthropodicus," an intracellular secondary endosymbiont from the hippoboscid louse fly Pseudolynchia canariensis. *Appl Environ Microbiol*, 72:2997-3004.

Dale, C., T. Jones, and M. Pontes. 2005. Degenerative evolution and functional diversification of type-III secretion systems in the insect endosymbiont Sodalis glossinidius. *Mol Biol Evol*, 22:758-66.

Dale, C., and I. Maudlin. 1999. Sodalis gen. nov. and Sodalis glossinidius sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly Glossina morsitans morsitans. *Int J Syst Bacteriol*, 49 Pt 1:267-75.

Dale, C., and N. A. Moran. 2006. Molecular interactions between bacterial symbionts and their hosts. *Cell*, 126:453-65.

Danneels, E. L., D. B. Rivers, and D. C. de Graaf. 2010. Venom Proteins of the Parasitoid Wasp Nasonia vitripennis: Recent Discovery of an Untapped Pharmacopee. *Toxins (Basel)*, 2:494-516.

Darby, A. C., J. H. Choi, T. Wilkes, M. A. Hughes, J. H. Werren, G. D. Hurst, and J. K. Colbourne. 2010. Characteristics of the genome of Arsenophonus nasoniae, son-killer bacterium of the wasp Nasonia. *Insect Mol Biol*, 19 Suppl 1:75-89.

Darling, A. C. E., B. Mau, F. R. Blattner, and N. T. Perna. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*, 14:1394-403.

Darling, A. E., B. Mau, and N. T. Perna. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*, 5:e11147.

Darling, A. E., I. Miklos, and M. A. Ragan. 2008. Dynamics of genome rearrangement in bacterial populations. *PLoS Genet*, 4:e1000128.

Derzelle, S., S. Ngo, E. Turlin, E. Duchaud, A. Namane, F. Kunst, A. Danchin, P. Bertin, and J. F. Charles. 2004a. AstR-AstS, a new two-component signal transduction system, mediates swarming, adaptation to stationary phase and phenotypic variation in Photorhabdus luminescens. *Microbiology*, 150:897-910.

Derzelle, S., E. Turlin, E. Duchaud, S. Pages, F. Kunst, A. Givaudan, and A. Danchin.

2004b. The PhoP-PhoQ two-component regulatory system of Photorhabdus luminescens is essential for virulence in insects. *J Bacteriol*, 186:1270-9.

Dorel, C., P. Lejeune, and A. Rodrigue. 2006. The Cpx system of Escherichia coli, a strategic signaling pathway for confronting adverse conditions and for settling biofilm communities? *Res Microbiol*, 157:306-14.

do Vale, A., M. T. Silva, N. M. dos Santos, D. S. Nascimento, P. Reis-Rodrigues, C. Costa-Ramos, A. E. Ellis, and J. E. Azevedo. 2005. AIP56, a novel plasmid-encoded virulence factor of Photobacterium damselae subsp. piscicida with apoptogenic activity against sea bass macrophages and neutrophils. *Mol Microbiol*, 58:1025-38.

Duron, O., D. Bouchon, S. Boutin, L. Bellamy, L. Zhou, J. Engelstadter, and G. D. Hurst. 2008. The diversity of reproductive parasites among arthropods: Wolbachia do not walk alone. *BMC Biol*, 6:27.

Duron, O., T. E. Wilkes, and G. D. Hurst. 2010. Interspecific transmission of a male-killing bacterium on an ecological timescale. *Ecol Lett*, 13:1139-48.

Eguchi, Y., T. Okada, S. Minagawa, T. Oshima, H. Mori, K. Yamamoto, A. Ishihama, and R. Utsumi. 2004. Signal transduction cascade between EvgA/EvgS and PhoP/PhoQ two-component systems of Escherichia coli. *J Bacteriol*, 186:3006-14.

Ewing, B. and P. Green. 1998. Basecalling of automated sequencer traces using phred. II. Error probabilities. *Genome Res,* 8:186-194.

Ewing, B., Hillier, L., Wendl, M. and P. Green. 1998. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res,* 8:175-185.

Ferree, P. M., A. Avery, J. Azpurua, T. Wilkes, and J. H. Werren. 2008. A bacterium targets maternally inherited centrosomes to kill males in Nasonia. *Curr Biol*, 18:1409-14.

ffrench-Constant, R., and N. Waterfield. 2005. An ABC guide to the bacterial toxin complexes. *Adv Appl Microbiol*, 58C:169-183.

Fronzes, R., P. J. Christie, and G. Waksman. 2009. The structural biology of type IV secretion systems. *Nat Rev Microbiol*, 7:703-14.

Garcia-Calderon, C. B., J. Casadesus, and F. Ramos-Morales. 2007. Rcs and PhoPQ regulatory overlap in the control of Salmonella enterica virulence. *J Bacteriol*, 189:6635-44.

Georgellis, D., O. Kwon, and E. C. Lin. 2001. Quinones as the redox signal for the arc

two-component system of bacteria. *Science*, 292:2314-6.

Gherna, R. L., Werren, J.H., Weisburg, W., Cote, R., Woese, C.R., Mandelco, L., and D.J. Brenner. 1991. Arsenophonus nasoniae gen. nov., sp. nov., the causative agent of the son-killer trait in the parasitic wasp Nasonia vitripennis. *Int J Syst Bacteriol*, 41:563-565.

Gordon, D., Abajian, C. and P. Green. 1998. Consed: A graphical tool for sequence finishing. *Genome Res*, 8:195-202.

Grindle, N., J. J. Tyner, K. Clay, and C. Fuqua. 2003. Identification of Arsenophonus-type bacteria from the dog tick Dermacentor variabilis. *J Invertebr Pathol*, 83:264-6.

Groisman, E. A. 2001. The pleiotropic two-component regulatory system PhoP-PhoQ. *J Bacteriol*, 183:1835-42.

Heddi, A., A. M. Grenier, C. Khatchadourian, H. Charles, and P. Nardon. 1999. Four intracellular genomes direct weevil biology: nuclear, mitochondrial, principal endosymbiont, and Wolbachia. *Proc Natl Acad Sci U S A*, 96:6814-9.

Hinnebusch, B. J., A. E. Rudolph, P. Cherepanov, J. E. Dixon, T. G. Schwan, and A. Forsberg. 2002. Role of Yersinia murine toxin in survival of Yersinia pestis in the midgut of the flea vector. *Science*, 296:733-5.

Hueck, C. J. 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev*, 62:379-433.

Huger, A. M., S. W. Skinner, and J. H. Werren. 1985. Bacterial infections associated with the son-killer trait in the parasitoid wasp Nasonia (= Mormoniella) vitripennis (Hymenoptera: Pteromalidae). *J Invertebr Pathol*, 46:272-80.

Hypsa, V., and C. Dale. 1997. In vitro culture and phylogenetic analysis of "*Candidatus* Arsenophonus triatominarum," and intracellular bacterium from the triatomine bug, *Triatoma infestans*. *Int. J. Syst. Bacteriol.*, 47:1140-1144.

Kanehisa, M., and S. Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*, 28:27-30.

Krin, E., Chakroun, N., Turlin, E., Givaudan, A., Gaboriau, G., Bonne, I., Rousselle, J-C., Frangeul, L., Lacroix, C., Hullo, M-F., Marisa, L., Danchin, A., and S. Derzelle. 2006. Pleiotropic role of quorum-sensing autoinducer 2 in *Photorhabdus luminescens*. *Appl. Environ. Microbiol.*, 27:6439-6451.

Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra. 2009. Circos: an information aesthetic for comparative

genomics. *Genome Res*, 19:1639-45.

Lang, A. E., G. Schmidt, A. Schlosser, T. D. Hey, I. M. Larrinua, J. J. Sheets, H. G. Mannherz, and K. Aktories. 2010a. Photorhabdus luminescens toxins ADP-ribosylate actin and RhoA to force actin clustering. *Science*, 327:1139-42.

Lang, A. E., G. Schmidt, J. J. Sheets, and K. Aktories. 2010b. Targeting of the actin cytoskeleton by insecticidal toxins from Photorhabdus luminescens. *Naunyn Schmiedebergs Arch Pharmacol*, 383:227-35.

Loui, C., A. C. Chang, and S. Lu. 2009. Role of the ArcAB two-component system in the resistance of Escherichia coli to reactive oxygen stress. *BMC Microbiol*, 9:183.

Marshall, A.G. 1981. The ecology of ectoparasitic insects. Academic Press, London.

Moriya, Y., M. Itoh, S. Okuda, A. C. Yoshizawa, and M. Kanehisa. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*, 35:W182-5.

Nogge, G. 1981. Significance of symbionts for the maintenance of an optional nutritional state for successful reproduction in hematophagous arthropods. *Parasitol*, 82:101-104.

Novakova, E., and V. Hypsa. 2007. A new Sodalis lineage from bloodsucking fly Craterina melbae (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont Sodalis glossinidius. *FEMS Microbiol Lett*, 269:131-5.

Novakova, E., V. Hypsa, and N. A. Moran. 2009. Arsenophonus, an emerging clade of intracellular symbionts with a broad host distribution. *BMC Microbiol*, 9:143.

Pais, R., C. Lohs, Y. Wu, J. Wang, and S. Aksoy. 2008. The obligate mutualist Wigglesworthia glossinidia influences reproduction, digestion, and immunity processes of its host, the tsetse fly. *Appl Environ Microbiol*, 74:5965-74.

Pearson, M. M., M. Sebaihia, C. Churcher, M. A. Quail, A. S. Seshasayee, N. M. Luscombe, Z. Abdellah, C. Arrosmith, B. Atkin, T. Chillingworth, et al. 2008. Complete genome sequence of uropathogenic Proteus mirabilis, a master of both adherence and motility. *J Bacteriol*, 190:4027-37.

Pontes, M. H., M. Babst, R. Lochhead, K. Oakeson, K. Smith, and C. Dale. 2008. Quorum sensing primes the oxidative stress response in the insect endosymbiont, Sodalis glossinidius. *PLoS ONE*, 3:e3541.

Pontes, M. H., K. L. Smith, L. De Vooght, J. Van Den Abbeele, and C. Dale. 2011. Attenuation of the sensing capabilities of PhoQ in transition to obligate insect-bacterial association. *PLoS Genet*, 7:e1002349.

Raffa, R. G., and T. L. Raivio. 2002. A third envelope stress signal transduction pathway in Escherichia coli. *Mol Microbiol*, 45:1599-611.

Reading, N. C., D. A. Rasko, A. G. Torres, and V. Sperandio. 2009. The two-component system QseEF and the membrane protein QseG link adrenergic and stress sensing to bacterial pathogenesis. *Proc Natl Acad Sci U S A*, 106:5889-94.

Reitzer, L. 2003. Nitrogen assimilation and global regulation in Escherichia coli. *Annu Rev Microbiol*, 57:155-76.

Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*, 16:276-7.

Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell. 2000. Artemis: sequence visualization and annotation. *Bioinformatics*, 16:944-5.

Shaw, P. D., G. Ping, S. L. Daly, C. Cha, J. E. Cronan, Jr., K. L. Rinehart, and S. K. Farrand. 1997. Detecting and characterizing N-acyl-homoserine lactone signal molecules by thin-layer chromatography. *Proc Natl Acad Sci U S A*, 94:6036-41.

Shigenobu, S., H. Watanabe, M. Hattori, Y. Sakaki, and H. Ishikawa. 2000. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. *Nature*, 407:81-6.

Silva, M. T., N. M. Dos Santos, and A. do Vale. 2010. AIP56: A Novel Bacterial Apoptogenic Toxin. *Toxins (Basel)*, 2:905-18.

Stock, A. M., V. L. Robinson, and P. N. Goudreau. 2000. Two-component signal transduction. *Annu Rev Biochem*, 69:183-215.

Toh, H., B. L. Weiss, S. A. Perkin, A. Yamashita, K. Oshima, M. Hattori, and S. Aksoy. 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of Sodalis glossinidius in the tsetse host. *Genome Res*, 16:149-56.

Trowbridge, R. E., K. Dittmar, and M. F. Whiting. 2006. Identification and phylogenetic analysis of Arsenophonus- and Photorhabdus-type bacteria from adult Hippoboscidae and Streblidae (Hippoboscoidea). *J Invertebr Pathol*, 91:64-8.

Vendeville, A., K. Winzer, K. Heurlier, C. M. Tang, and K. R. Hardie. 2005. Making 'sense' of metabolism: autoinducer-2, LuxS and pathogenic bacteria. *Nat Rev Microbiol*, 3:383-96.

Vlisidou, I., I. Eleftherianos, S. Dorus, G. Yang, R. H. ffrench-Constant, S. E. Reynolds,

and N. R. Waterfield. 2010. The KdpD/KdpE two-component system of Photorhabdus asymbiotica promotes bacterial survival within M. sexta hemocytes. *J Invertebr Pathol*, 105:352-62.

Waterfield, N. R., D. J. Bowen, J. D. Fetherston, R. D. Perry, and R. H. ffrench-Constant. 2001. The tc genes of Photorhabdus: a growing family. *Trends Microbiol*, 9:185-91.

Waters, C. M., and B. L. Bassler. 2005. Quorum sensing: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol*, 21:319-46.

Watson, W. T., T. D. Minogue, D. L. Val, S. B. von Bodman, and M. E. Churchill. 2002. Structural basis and specificity of acyl-homoserine lactone signal production in bacterial quorum sensing. *Mol Cell*, 9:685-94.

Werren, J. H., S. W. Skinner, and A. M. Huger. 1986. Male-killing bacteria in a parasitic wasp. *Science*, 231:990-2.

Wilkes, T. E., A. C. Darby, J. H. Choi, J. K. Colbourne, J. H. Werren, and G. D. Hurst. 2010. The draft genome sequence of Arsenophonus nasoniae, son-killer bacterium of Nasonia vitripennis, reveals genes associated with virulence and symbiosis. *Insect Mol Biol*, 19 Suppl 1:59-73.

Wilkes, T. E., O. Duron, A. C. Darby, V. Hypsa, E. Novakova, and G. D. D. Hurst. 2011. The genus Arsenophonus. In E. Zchori-Fein, and K. Bourtzis (ed.), *Manipulative tenants: Bacteria associated with arthropods*. CRC Press, Boca Raton, FL.

Williams, P. 2007. Quorum sensing, communication and cross-kingdom signalling in the bacterial world. *Microbiology*, 153:3923-38.

Xavier, K. B., and B. L. Bassler. 2003. LuxS quorum sensing: more than just a numbers game. *Curr Opin Microbiol*, 6:191-7.

Yuan, J., B. Wei, M. Shi, and H. Gao. 2011. Functional assessment of EnvZ/OmpR two-component system in Shewanella oneidensis. *PLoS One*, 6:e23701.

CHAPTER 4

REPLACEMENT OF NATIVE SYMBIONTS IN THE HIPPOBOSCID LOUSE FLY, PSEUDOLYNCHIA CANARIENSIS

Abstract

The pigeon louse fly, *Pseudolynchia canariensis*, harbors two distinct bacterial symbionts that are maternally transmitted to larvae *in utero*: a bacteriome-associated primary symbiont related to the tsetse fly primary symbiont, *Wigglesworthia glossinidia*, and a secondary symbiont, *Candidatus* Arsenophonus arthropodicus, which is distributed throughout multiple insect host tissues. In the current study, we show that genetically modified *Ca.* A. arthropodicus can initiate infections in louse flies after microinjection into adult flies and fly puparia. Vertical transmission of the mutant strain was observed only after injection into flies during the pupal stage, and in many insects this led to a complete replacement of native wild type symbionts with the mutant strain, although complete loss of *Ca.* A. arthropodicus symbionts was also observed in some cases. Changes in the population structure of the symbiont *Ca.* A. arthropodicus within the louse fly did not affect the existing population of primary symbionts, which were always retained, regardless of whether the populations of *Ca.* A. arthropodicus were replaced or lost. The ability to infect louse flies with genetically modified symbionts and maintain infections over multiple generations may prove very useful in facilitating investigation

into symbiont gene function and pursuing strategies of paratransgenesis in viviparous insects such as tsetse flies.

Introduction

Many insects maintain long-term associations with bacterial symbionts that enable them to persist in a specialized niche (Buchner, 1965; Feldhaar, 2011). Our knowledge of insect-bacterial symbiotic systems has expanded greatly over the past decade, due in large part to the generation of whole genome sequences for many of these bacteria (Moran et al., 2008). One general feature that has emerged is that the genome sizes of obligate symbionts are substantially reduced in comparison to free-living relatives (McCutcheon and Moran, 2012). The genomes of insect symbionts are known to degenerate over time as a result of relaxed selection, which mediates the loss of genes that do not provide critical functions in the isolated and static environment of the insect host (Dale and Moran, 2006; Moran and Plague, 2004). In addition, host restriction deprives symbionts of the opportunity to undertake lateral gene transfer with other bacteria, catalyzing a degenerative evolutionary trajectory in which bacteria retain only those genes that are vital in the symbiotic lifestyle (Moran and Plague, 2004). Consequently, characterization of the reduced gene inventories of these insect symbionts has provided important insight into their metabolic capabilities and functions. For example, the genome of *Buchnera aphidicola,* an ancient symbiont of the pea aphid, is only 640 kb in size (Shigenobu et al., 2000), greatly reduced from that of a close free-living relative, *Escherichia coli* (4.6 Mb) (Blattner et al., 1997), indicating it has undergone extensive gene loss. The gene inventory of *B. aphidicola* retains only those genes necessary for its symbiotic lifestyle, including pathways for the production of

essential amino acids that are lacking in the aphid diet of plant sap, supporting the idea that this symbiont plays a role in host diet supplementation (Shigenobu et al., 2000; Douglas, 1998).

While reduced gene inventories have proved useful in predicting the functional capabilities of insect symbionts, little is known currently about the basic molecular mechanisms underlying interactions between insects and their symbiotic bacteria. This is because only a small number of insect symbionts have been cultured in the laboratory, all of which are predicted to have a relatively recent origin of association with their insect hosts (Gherna et al, 1991; Hypsa and Dale, 1997; Dale et al, 1999; Dale et al, 2006; Darby et al, 2005; Sabri et al, 2010). Because of this recent association, the genomes of these symbionts are still in the early stages of degeneration and have not yet undergone extensive gene loss that is common in ancient symbionts. For example, the genome size of *S. glossinidius,* a cultured symbiont of tsetse flies, is 4.29 Mb (Toh et al, 2006), similar in size to *E. coli*. However, its genome is in the process of degeneration, as evidenced by the fact that many of the genes in the *S. glossinidius* genome have acquired nonsense mutations or frameshifts that render them non-functional, leading to a reduced overall coding capacity of only 50% (Toh et al, 2006; Belder et al, 2010).

In addition to culturing, methods have been developed to genetically modify certain recently derived insect symbionts. Much of this work has focused on *S. glossinidius*, which is readily amenable to transformation with plasmid DNA (Beard et al., 1993; Cheng and Aksoy, 1999; Weiss et al., 2006). However, exogenous plasmid transformations are known to lack long-term stability in the absence of antibiotic selection (De Gelder, 2007), so recent efforts have focused on engineering symbionts

with stable chromosomal modifications. Most recently, the bacteriophage lambda Red recombineering strategy (Datsenko and Wanner, 2000) was adapted to engineer targeted chromosomal gene replacements in *S. glossinidius* (Pontes and Dale, 2011; Pontes et al., 2011). In one of these studies, native tsetse fly hosts as well as a novel louse fly host, *Pseudolynchia canariensis,* were superinfected with a *S. glossinidius phoP* mutant to determine its ability to infect insects (Pontes et al., 2011). The PhoP mutant strain was found to be more sensitive to antimicrobial peptides and was ineffective in colonizing both of the insect hosts, indicating that PhoP plays a crucial role in mediating resistance to the insect immune system and allowing *S. glossinidius* to establish and/or maintain an infection.

While the phenotypic effect of *phoP* in *S. glossinidius* was relatively easy to ascertain using a superinfection approach, the effects of more subtle mutations might only be understood in long-term infection experiments in which native symbionts have been replaced with genetically modified variants *in vivo*. However, these approaches have proven particularly challenging in insects that maintain more than one type of mutualistic endosymbiont. For example, in addition to *S. glossinidius*, tsetse flies maintain an ancient primary symbiont called *Wigglesworthia glossinidia* that is essential for insect survival and reproduction (Aksoy, 1995). Thus, to assess the phenotypic effect of a recombinant *S. glossinidius* strain, one must selectively remove a resident (native) population of *S. glossinidius* from tsetse flies without affecting *W. glossinidia*. In addition to exploring the functions of symbiont genes, the aforementioned symbiont replacement strategy is important in the context of paratransgenesis, which involves using symbionts to express transgenes in insects to reduce the capability of the insect to

transmit disease (Durvasula et al, 1999; Beard, et al., 1998). This is of particular relevance to *S. glossinidius*, because its tsetse fly host is an important vector of trypanosomes that cause sleeping sickness in humans and nagana in animals in sub-Saharan Africa (Aksoy et al, 2001; Aksoy et al, 2008).

In the current study we adapted the lambda Red recombineering strategy (Datsenko and Wanner, 2000) to engineer a chromosomally modified strain of *Candidatus* Arsenophonus arthropodicus, the secondary symbiont of the pigeon louse fly, *Pseudolynchia canariensis* (Dale et al., 2006). This symbiont is a member of the *Arsenophonus* group of closely related bacteria that have been identified in a broad range of distantly related insect hosts with diverse lifestyles (Dale et al., 2006; Novakova et al., 2009). Using the recombineering strategy, we engineered a "phenotypically neutral" strain of *Ca.* A. arthropodicus, in which a predicted pseudogene was interrupted with an antibiotic resistance marker. We used this recombinant strain to investigate the ability of this symbiont to initiate and maintain an infection in louse flies after introduction by microinjection, as well as the ability of the recombinant symbionts to undergo vertical transmission to the next generation. We then developed a pupal microinjection strategy that facilitated complete replacement of native *Ca.* A. arthropodicus in its natural insect host, without disrupting an existing population of primary symbionts.

Materials and Methods

Insect and bacterial cultures

Adult *P. canariensis* flies were maintained on rock pigeons (*Columbia livia*) in the University of Utah Animal Care Facility. Fly puparia were collected and incubated at 28°C with 65% humidity until fly emergence. Cultures of *Ca.* A. arthropodicus were

maintained at 28°C in liquid Mitsuhashi and Maramorosch medium (MM medium) as described previously (Dale et al., 2006) or in MM medium without glucose (-Glu) as indicated. Rapidly dividing cultures for transformation were obtained by shaking liquid cultures in a water bath at 28°C, 150 rpm, following an initial two day growth period in a non-shaking incubator. For bacterial clone isolation, liquid cultures were plated on 1.5% MM agar plates supplemented with 5% defibrinated horse blood and incubated at 28°C under microaerophilic conditions (5% O_2, 10% CO_2, balanced with N_2) for 5-7 days. Media supplements of 25 μg/ml kanamycin (KM), 12.5 μg/ml chloramphenicol (CM), and 50 μg/ml polymyxin B (PB) were added as appropriate.

Generation of gene replacement alleles

Gene replacement alleles containing the kanamycin resistant marker derived from the pCR4-TOPO vector (Invitrogen) flanked by 600 bp of sequence identical to the regions targeted for homologous recombination, flanking the *adhE* pseudogene, were generated according to a three-step PCR protocol (Derbise et al., 2003) using AccuTaq LA DNA polymerase (Sigma-Aldrich). First, 600 bp upstream and downstream regions were amplified using the *adhE*-UpR and *adhE*-DnF primers with 20 bases of sequence matching the ends of the kanamycin resistance allele (Table 4.1). These PCR products were combined in equimolar ratios with an amplicon of the kanamycin resistance allele and a fusion construct was amplified using the *adhE*-UpF and *adhE*-DnR primers. This PCR was then repeated on the product obtained to enhance the amount of full-length product generated, which was then purified and concentrated using a Microcon centrifugal filter (Millipore). PCR reactions were performed in 50 μl volumes containing 10 mM of each dNTP, 2.5 mM $MgCl^2$, 20 pmol/μl each primer, 2% DMSO and 0.05

Table 4.1. Sequences of primers used for λ Red recombineering, qPCR and PCR.

Usage	Primer Name	Sequence (5' --> 3')
Lambda Red:	KmTopo_F	TAGACTGGGCGGTTTTATGG
	KmTopo_R	GGGAGTCAGGCAACTATGGA
	AdhE_Up_F	TTGCCACAGAAAAATTAGCC
	AdhE_Up_R	CCATAAAACCGCCCAGTCTAACACGCACGACAAGTTCATT
	AdhE_Dn_F	TCCATAGTTGCCTGACTCCCCCAATATGATCGCCCTCAAG
	AdhE_Dn_R	ATAGAGGCGGGTGAAAAACC
	AdhE_Out_F	CGCCAGCAATACGAAAAGA
	AdhE_Out_R	ATGAACGGGGAATGAATACG
qPCR:	Rec_Km_F	GGCACCTATCTCAGCGATCT
	Rec_Km_R	TTCAATTTTGTTAGCTGTGCG
	AdhE_WT_F	AAAACCTGCCATCGGTGTAG
	AdhE_WT_R	CTCTGACGCGCAAATAACAC
Primary screen:	Prim_F	AACGGGGAAGCTATGCTTCTGC
	Prim_R	GAGGTTTGCTAACTTTTGCAAGCT

units/μl polymerase. PCR amplifications were run on a BioRad iCycler with the following cycling conditions: initial denaturation at 98°C for 30 sec, followed by 35 cycles of denaturation at 98°C for 15 sec, annealing at 60°C for 20 sec, and elongation at 68° for 1 min/kb, followed by a final extension step at 72° for 7 min.

Transformation of Ca. A. arthropodicus with pKD46-CM

Ca. A. arthropodicus was transformed by electroporation with plasmid pKD46-CM, which contains the lambda Red recombineering genes and a chloramphenicol resistance marker. Briefly, 100 ml of cells were harvested by centrifugation at 6000 x *g* for 10 min at 4°C. Cells were washed three times with ice-cold 10% glycerol and concentrated 200-fold. 125 μl aliquots of cells were transformed by electroporation in a 1 mm cuvette at 1800 V/cm. Cells were immediately transferred to 50 ml tubes containing 20 ml MM (-Glu) and allowed to recover overnight at 28°C with shaking.

Cells were then harvested by centrifugation at 6000 x *g* for 10 min and resuspended in 200 μl MM and plated on MM blood agar plates (CM) to select for transformants.

Lambda Red recombineering

Ca. A. arthropodicus harboring plasmid pKD46-CM was grown in liquid MM (-Glu, +CM) to mid-log phase in a shaking water bath at 110 rpm, 28°C. Expression of lambda Red genes from pKD46-CM was induced with 0.2% arabinose for 8 hours. Cells were then transformed as described above using 5 μg of the replacement construct and plated on MM blood agar plates (CM/KM) to select for recombinants. Plates were incubated for 7 days at 28°C under microaerophilic conditions. A recombinant clone that was verified by PCR and sequencing was then inoculated into 20 ml MM without selection for pKD46-CM and subcultured every 3 days for a total of five passages, to obtain recombinants that had lost the lambda Red plasmid pKD46-CM. Cells were seeded onto MM blood agar plates (KM) and replica plated after growth to identify clones that were sensitive to chloramphenicol. One CM sensitive clone was selected and used for subsequent microinjections.

Microinjection of symbionts into insects

Bacteria were grown to mid-log phase and cells were collected in a 1.5 ml tube by centrifugation at 6000 x *g* for 5 min. Cells were washed and resuspended in 1X PBS and diluted to an OD_{600} of 0.01. Then, 0.1 μl of cell suspension (~$5x10^4$ cells) was injected into adult louse flies in the ventral thorax within 24 hours of emergence from the puparia, or into puparia at the posterior end above the spiracles within 24 hours of deposition, before the pupal case had fully hardened. Flies were then maintained on pigeons and

their offspring were collected at various timepoints following injection and frozen at -80°C for subsequent DNA extraction.

PCR and real-time quantitative PCR

Total insect DNA was extracted from louse flies using the DNeasy Blood and Tissue Kit (Qiagen) according to manufacturer's instructions. The isolated DNA was screened for the presence of symbionts by PCR using primers targeting the resistance cassette and flanking regions for *Ca.* A. arthropodicus, or the 16S rDNA of the primary symbiont (Table 4.1). Quantitative PCR (qPCR) was performed in triplicate on insect DNA samples in 20 μl reactions using Fermentas Maxima SYBR Green/Fluorescein qPCR 2X Master Mix containing 0.5 μM primers and 5 μl template DNA and run on a BioRad iCycler iQ Real-Time PCR Detection System. Cycling conditions consisted of one cycle at 95°C for 10 min, followed by 40 cycles consisting of 95°C for 15 sec and 60°C for 1 min, followed by 1 min at 95°C and 1 min at 55°C, and finishing with a melt curve during which temperatures were increased by 0.5°C per cycle for 80 cycles, from 55°C to 95°C. Relative population ratios were estimated against a single standard curve constructed by cloning target PCR products for both the WT and Δψ*adhE* strain into a plasmid vector. This plasmid was linearized, purified and diluted to a concentration of $2x10^6$ molecules/μl and 10-fold serial dilutions were prepared and used for standard curve calculations. All qPCR assays had PCR efficiencies between 90 - 100% and a melt curve was performed at the end of each run to ensure that only the desired product was amplified. For statistical analyses, the non-parametric Mann-Whitney test was used to determine if there were significant differences between groups.

Colony counts from puparia

For colony counts, puparia were surface sterilized once in 6% sodium hypochlorite for 5 min, followed by a 70% ethanol rinse, three water washes and a final rinse in MM media. Puparia were then individually homogenized in 1.5 ml tubes in 100 μl of MM media. Ten-fold serial dilutions were prepared and 100 μl were spread onto MM-blood-agar plates (PB). After growth on nonselective media, colonies were transferred to fresh plates (KM) to determine the ratios of WT to kanamycin-resistant cells.

Results

Construction of a "phenotypically neutral" strain of Ca. A. arthropodicus

To develop a microinjection strategy for introducing genetically modified *Ca.* A. arthropodicus into louse flies, it was necessary to engineer a symbiont strain that retained the same phenotypic characteristics as the wild type (WT) strain, but would be distinguishable from the native WT strain already resident in the fly. Using the lambda Red recombineering procedure (Derbise et al., 2006; Pontes and Dale, 2011), we generated a genetically distinct, but phenotypically neutral strain of *Ca.* A. arthropodicus (Δψ*adhE*), by interrupting a chromosomal pseudogene, ψ*adhE*, with a kanamycin resistance cassette (Fig. 4.1A). We verified the genetic status of the mutant strain by PCR (Fig. 4.1B) using primers that flanked the homologous recombination site (Table 4.1). PCR products obtained using primers located in the Δψ*adhE* flanking DNA were sequenced to confirm the integrity of the chromosomal modification.

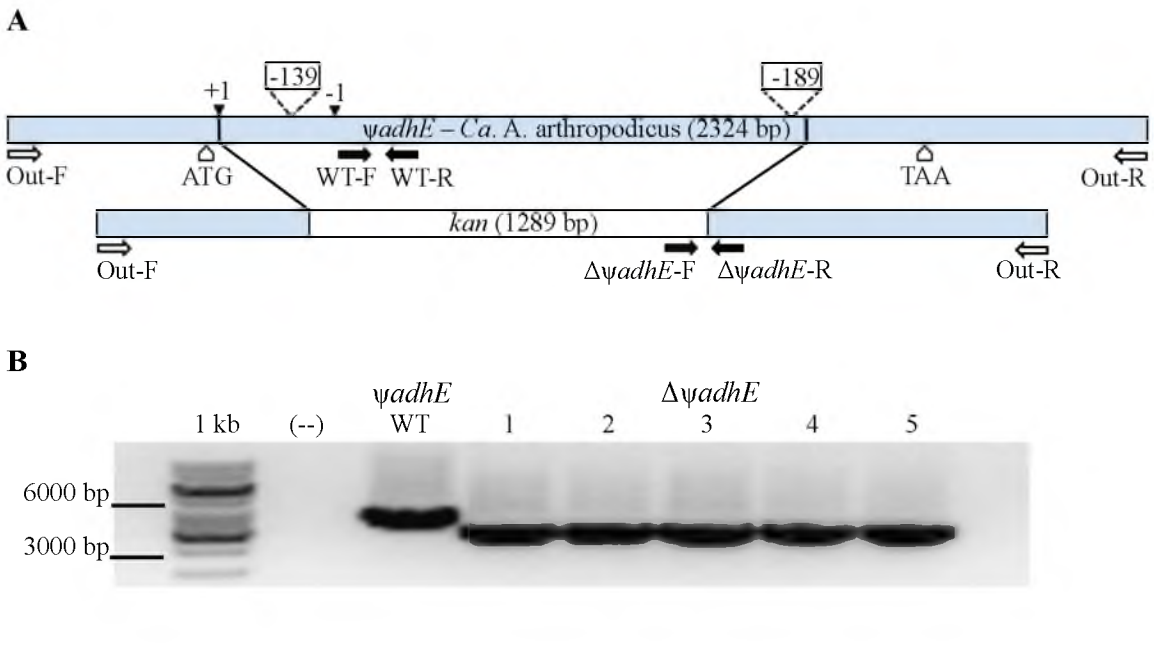


Figure 4.1. Locations of mutations in pseudogene ψ*adhE* of *Ca.* A. arthropodicus and the resulting Δψ*adhE* strain after lambda-Red mediated homologous recombination. A. Predicted mutations present in pseudogene ψ*adhE* of *Ca.* A. arthropodicus (2,324 bp) deduced from an alignment with an intact copy of *adhE* from *Arsenophonus nasoniae* (2,667 bp). Black triangles indicate a single base pair frameshifting insertion or deletion. White boxes with dotted lines designate deletions in the *Ca.* A. arthropodicus ψ*adhE* gene of the indicated size. Solid black lines denote the site at which homologous recombination occurred, replacing most of the ψ*adhE* gene with a kanamycin resistance cassette (*kan*) to create strain Δψ*adhE*. Black arrows indicate primer sets used to differentiate WT and Δψ*adhE* strains for quantitative PCR (qPCR). B. PCR amplification of the *adhE* region in WT and 5 Δψ*adhE* clones, using primers that bind outside of the recombination sites, indicated by white arrows in (A).

*The Δψadh*E *strain displays the same growth rate in vitro as the WT*

To determine if the Δψ*adhE* strain maintained a growth rate *in vitro* that was similar to that of the WT, we inoculated each strain into fresh media and measured the optical density (OD_{600}) over the course of growth. Both strains displayed near identical growth rates (Fig. 4.2A). To confirm that the two strains were equally competitive *in vitro*, equivalent amounts of the Δψ*adhE* and WT strains were combined and inoculated

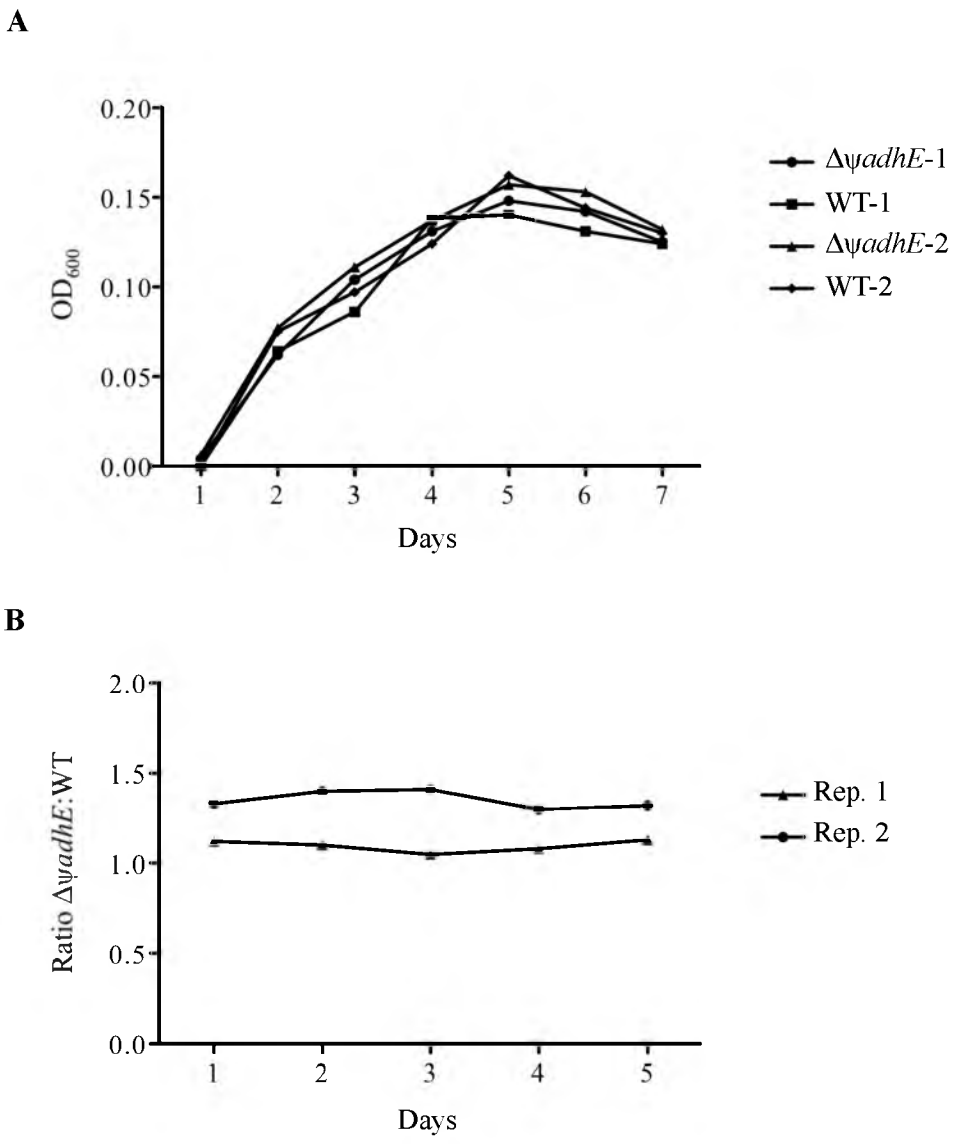


Figure 4.2. Growth curves of Δψ*adhE* and WT strains and competitive growth assays. A. Growth curves of Δψ*adhE* and WT strains in MM media, performed in duplicate. B. Ratio of Δψ*adhE*:WT strains over time during competitive growth *in vitro* in MM media. Two replicates were analyzed by qPCR.

into fresh media and sampled daily. The numbers of cells of each strain were then quantified by qPCR. The ratio of Δψ*adhE*:WT cells remained constant over the course of growth (Fig. 4.2B).

The Ca. A. arthropodicus population increases throughout the fly life cycle

The louse fly has an unusual life cycle similar to that of the tsetse fly, in which reproduction occurs by adenotrophic viviparity, wherein a single egg is fertilized during each gonotrophic cycle and is maintained by the female through the larval stages *in utero* (Leak, 1999; Bequaert, 1953). During development, the larva obtains nourishment and is concurrently inoculated with its symbiotic flora via feeding on milk gland secretions (Cheng and Aksoy, 1999; Attardo et al., 2008) until larviposition occurs, just prior to pupation. After the pupal stage, a teneral louse fly emerges and reaches reproductive maturity within a week (Bequaert, 1953). To determine the population density of *Ca.* A. arthropodicus throughout louse fly development, we used qPCR to measure the number of genome copies of native wild type *Ca.* A. arthropodicus over the course of development. As shown in Figure 4.3, the symbiont population was at its lowest levels in the early larval stages, indicating that the mother only transmits a small proportion of her symbiont population to the developing larva. The symbiont population increased rapidly from the early larval to mid pupal stages, and underwent a significant increase from the teneral stage to reach maximal levels in reproductively mature adults (Fig. 4.3).

Natural infection rate of louse flies with Ca. A. arthropodicus

To determine if *Ca.* A. arthropodicus is naturally transmitted to all offspring maternally, we screened 100 louse fly puparia by PCR, using primers specific to the wild

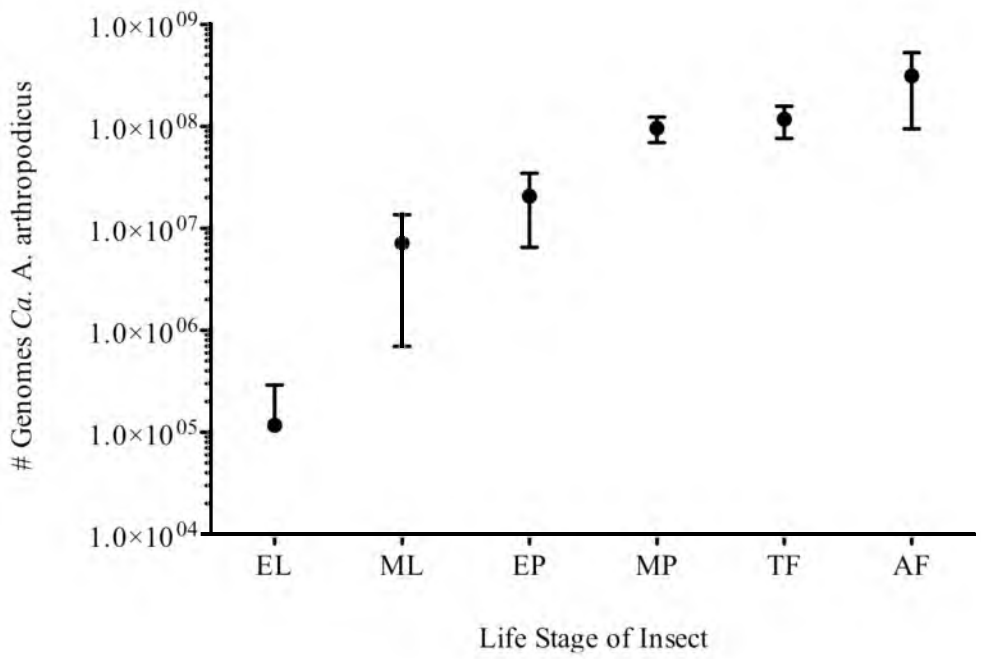


Figure 4.3. Number of *Ca.* A. arthropodicus symbiont genomes throughout host development. The WT symbiont population was measured by qPCR for 10 individuals in each of the following insect stages: early larval stage (EL), mid larval stage (ML), early pupal stage (EP), mid pupal stage (MP), teneral female fly (TF) and reproductively mature adult female fly (AF). The mean is indicated, with error bars representing standard deviations. The differences between each life stage were significant ($p<0.05$) except between the MP and TF stages ($p=0.2882$), using Fisher's t-test between successive life stages.

type strain (Table 4.1). *Candidatus* Arsenophonus arthropodicus was detected in all 100 puparia, indicating that it has a natural transmission rate of 100%.

Superinfected adult louse flies maintain but fail to transmit Δψ adhE symbionts to their offspring

We first attempted to superinfect adult louse flies by injecting Δψ*adhE Ca.* A. arthropodicus into the ventral thorax of adult flies just after their emergence from the pupal case. We then used qPCR to quantify the population levels of both the Δψ*adhE*

and WT strains in the adult flies eight days after microinjection, when the flies were fully mature and began to reproduce, to determine if the introduced bacteria were able to establish an infection. Both the Δψ*adhE* and WT strains were detected in all of the microinjected flies in similar quantities (Fig. 4.4A). There was no significant difference in the average population numbers of the Δψ*adhE* strain compared to the WT strain in the injected flies (p=0.1655). However, the Δψ*adhE* strain was detected at a very low level of infection in only 2/16 of the F1 offspring derived from microinjected adult flies (Fig. 4.4B), indicating it was not efficiently transmitted to offspring. Furthermore, in one of the F1 flies, neither the Δψ*adhE* nor the WT strain of *Ca.* A. arthropodicus was detected.

*Microinjection of early louse fly puparia led to replacement of WT symbionts with Δψ*adhE* symbionts in many offspring.*

Based on the symbiont population profile, we predicted that injecting the Δψ*adhE* symbionts earlier in the fly life cycle would yield higher initial ratios of Δψ*adhE*:WT bacteria, which might increase the probability that the Δψ*adhE* strain would undergo vertical transmission. To obtain a higher initial ratio of Δψ*adhE* to wild type symbionts, we microinjected Δψ*adhE Ca.* A. arthropodicus into louse fly puparia collected immediately after deposition by adult females, when the native symbiont population is at a low level relative to teneral and mature flies (Fig, 4.3). We collected teneral flies that emerged from injected puparia and examined the population ratios of the Δψ*adhE* and WT strains for a subset of the flies (Fig. 4.5), allowing the remaining flies to reproduce. Following injection of puparia, emerged teneral flies showed a mixed population of Δψ*adhE* and WT symbionts (Fig. 4.5), as was also observed in the injected adults (Fig.

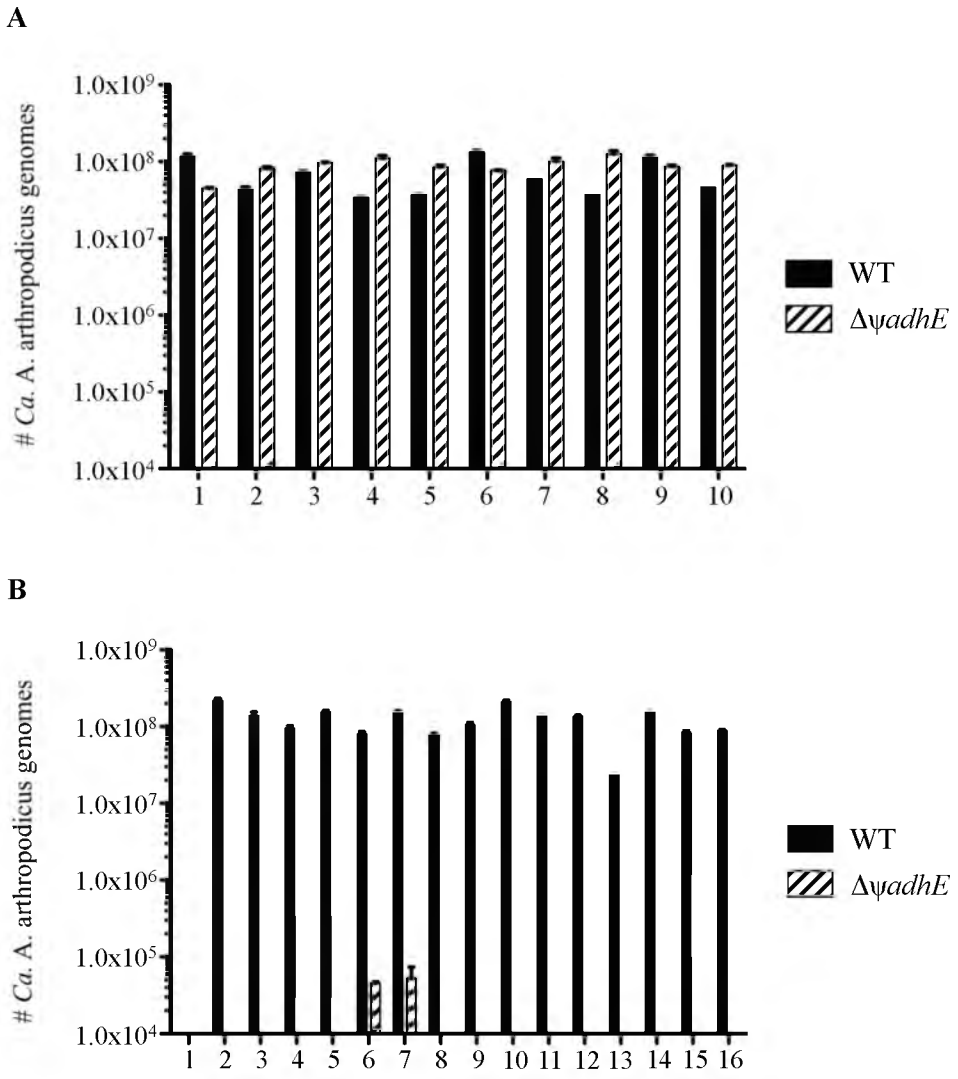


Figure 4.4. Number of *Ca.* A. arthropodicus symbiont genomes in microinjected adult flies and F1 offspring. A. qPCR analysis of population levels of the WT and Δψ*adhE* strains in individual adult flies 8 days postmicroinjection. Average populations of each strain were not significantly different (p=0.1655, Mann-Whitney test). B. qPCR analysis of F1 offspring of microinjected adult flies. Error bars represent standard deviations.

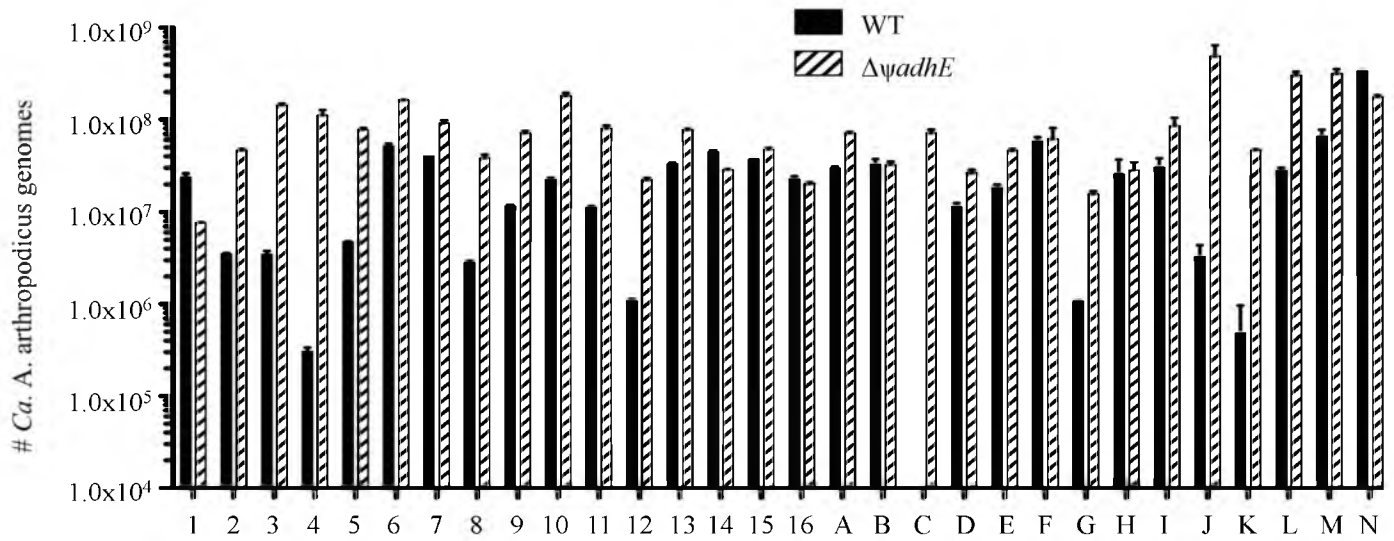


Figure 4.5. Number of *Ca.* A. arthropodicus symbiont genomes in two sets of microinjected puparia. qPCR analysis of population levels of the WT and Δψ*adhE* strains in individual teneral flies that were injected as early puparia. Results from two separate replicates are shown, with individuals derived from the first replicate labeled with numbers, and individuals from the second replicate labeled with letters. Error bars represent standard deviations.

4.4A). However, we detected significantly higher ratios of the Δψ*adhE* strain relative to the WT strain in comparison to insects injected as tenerals ($p<0.0001$).

The F1 offspring derived from these microinjected puparia showed a mixture of symbiont population types, but a majority of the F1 puparia (63%) were found to maintain only the Δψ*adhE* strain (Fig. 4.6A), indicating that not only was the introduced strain vertically transmitted, but a complete replacement of the WT symbionts occurred over the first generation in many lines. The 37% of F1 offspring that did not show complete replacement of the WT symbionts harbored symbiont populations composed of either WT symbionts alone (7%), a mixed population of Δψ*adhE* and WT symbionts (20%), or no *Ca.* A. arthropodicus (10%; Fig. 4.6A).

In the F2 generation, even fewer flies maintained WT symbionts as evidenced by our qPCR assays (Fig. 4.6B). Half of the flies harbored an infection of the Δψ*adhE* strain alone, while 7% retained only the WT strain and 43% were aposymbiotic, with no mixed populations observed (Fig. 4.6B). Both male and female representatives were identified for all of the different symbiont population structures observed, indicating that these effects were not sex-specific.

To investigate whether the native primary symbiont of the louse fly was also affected by the microinjection procedure, we screened fly samples by PCR, using primers that specifically target the 16S rDNA of the primary symbiont. In all offspring derived from the injected puparia, the primary symbiont was present, even when *Ca.* A. arthropodicus was absent (Fig. 4.6A,B), indicating that the primary symbiont was undisturbed by the microinjection procedure.

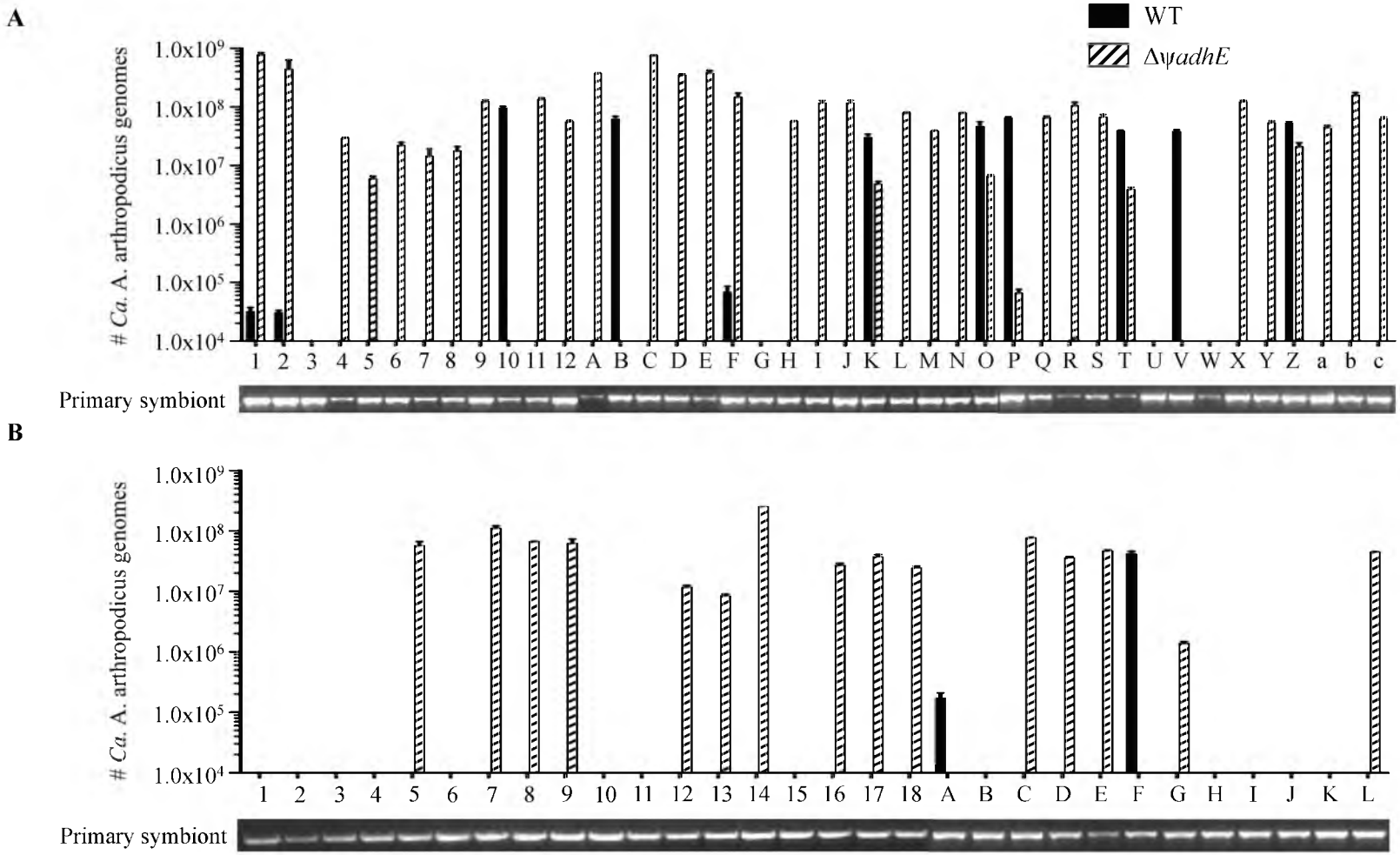


Figure 4.6. Number of *Ca.* A. arthropodicus symbiont genomes in F1 and F2 offspring of microinjected puparia. A. qPCR analysis of populations of the WT and Δψ*adhE* strains in F1 offspring. B. qPCR analysis of F2 generation. Error bars represent standard deviations. Gel images below graphs show PCR amplicons derived from primers targeting the louse fly primary symbiont 16S rDNA.

Plating experiments validated the replacement and loss of Ca. A. arthropodicus in F2 puparia.

A subset of F2 puparia (#13-18 in Fig. 4.6B) were subjected to culture on agar plates prior to DNA extraction to verify the identity of *Ca.* A. arthropodicus strains residing in these insects. One pupal extract (#15) had no growth, indicating that neither the WT nor Δψ*adhE* strain of *Ca.* A. arthropodicus was present, while the remaining five pupal extracts produced colonies on non-selective media, demonstrating the presence of *Ca.* A. arthropodicus. Of these five F2 puparia, all of the colonies examined displayed resistance to kanamycin (Table 4.2), indicating that the *Ca.* A. arthropodicus population in these insects consisted solely of the Δψ*adhE* strain. These results are entirely consistent with those obtained from the qPCR assays (Fig. 4.6B).

Discussion

In the current study, we investigated the ability of *Ca.* A. arthropodicus to initiate infections in the pigeon louse fly, *P. canariensis*, after microinjection into two different fly life cycle stages. The generation of a stable, genetically modified variant of *Ca.* A. arthropodicus strain (Δψ*adhE*), that maintains the same phenotypic characteristics as the wild type strain, enabled us to superinfect louse flies harboring a native wild-type population of symbionts and differentiate between the two strains throughout the course of infection. Our results show that louse flies superinfected with Δψ*adhE Ca.* A. arthropodicus can sustain an infection after microinjection into adult flies, but vertical transmission of the Δψ*adhE* strain to offspring only occurred in a small percentage of flies. Given that the modified symbionts were able to replicate to population numbers

Table 4.2. Number of colonies from F_2 pupal extracts demonstrating kanamycin resistance.

F_2 sample	Total colonies	Km-resistant
13	386	386
14	251	251
15	0	0
16	600	600
17	800	800
18	1000	1000

approximately equivalent to those of the WT strain, one might expect to observe transmission of the Δψ*adhE* symbionts in addition to the WT to offspring of injected flies if transmission is based strictly on symbiont populations sizes within flies. However, the absence of the Δψ*adhE* strain in most offspring of flies that were microinjected at the teneral stage indicates that there are additional factors beyond population density that play important roles in the allocation of symbionts to the next generation. It is possible that the microinjected symbionts were unable to initiate successful infections in the tissues required to undergo vertical transmission, such as the milk glands. If the Δψ*adhE* symbionts were unable to reach the proper tissues from their location of introduction or if these target tissues were already colonized by WT symbionts, the Δψ*adhE* strain may have been unable to create the local population density necessary to observe transmission to offspring.

Transmission to louse fly offspring was observed consistently following injection into freshly deposited puparia and, in many cases, this led to a complete replacement of the WT symbionts with the Δψ*adhE* strain. This may be due in part to the population dynamics of vertically transmitted symbionts, in which symbiont population bottlenecks occur each generation during insect reproduction, when only a small proportion of the

parental symbiont population is transmitted to offspring (Moran, 1996; Moran et al., 2008; Rio et al., 2006). Our results show that the louse fly symbiont population increases throughout the fly life cycle, with lowest levels observed in larvae and highest levels observed in reproductively mature adults. Therefore, by injecting earlier in the life cycle, we introduced a higher initial proportion of Δψ*adhE*:WT symbionts, which likely sustained increased ratios of Δψ*adhE*:WT into adulthood. The population bottleneck during reproduction may have then served to fix the Δψ*adhE* strain in the offspring. However, we cannot rule out the possibility that microinjected symbionts had higher viability or were preferentially allocated to tissues that facilitated enhanced transmission. Such preferential allocation may have resulted from the introduction of microinjected symbionts into a different location in the body rather than the natural route of infection through the fly gut. This variation might have led to different patterns of tissue colonization by the two symbiont strains, resulting in favored transmission of the injected strain over the native strain, the opposite result from that observed after microinjection into teneral flies.

Additionally, introduction of symbionts in a different manner may have led to a host response that played a role in the symbiont replacements or symbiont losses observed. By artificially increasing the total number of secondary symbionts present in the fly in a specific tissue location, a response might have been triggered in the host to keep the symbiont population within a certain density range or tissue. It is predicted that in symbiosis, the host maintains an ability to control symbiont numbers by imposing "sanctions" on symbionts that attempt to infect additional tissues or grow to a level that is harmful to the host (Reynolds and Rolff, 2008). Recent studies in weevils have shown

that expression of various antimicrobial peptides plays an important role in maintaining and controlling the association with its bacteriome-associated primary symbiont (Anselme et al., 2008; Login et al., 2011). For example, Login et al. (2011) showed that expression of coleoptericin-A (ColA) by weevils within their bacteriomes is involved in regulating symbiont population numbers by inhibiting bacterial cell division. Inhibiting ColA function in weevils led to smaller symbionts that invaded additional tissues outside of the bacteriome (Login et al., 2011). The presence of symbionts outside of the bacteriome in weevils has been shown to lead to a host immune response, with the induction of a range of other antimicrobial peptides, presumably because symbionts outside of their expected tissue distribution are instead identified as intruders (Anselme et al., 2008). Thus, the injection of an additional load of symbionts beyond the natural inoculant into louse fly pupae or hemolymph might lead to a host response that eliminates or decreases the overall symbiont population or tissue-specific populations to such an extent that symbionts are not efficiently transmitted. Artificially inflated symbiont levels due to microinjection might persist for more than one generation, so perhaps loss of symbionts can occur until the symbiont population stabilizes to an appropriate level. It will be useful to examine additional generations to determine if the number of aposymbiotic flies continues to increase or eventually stabilizes, and whether this relates to the total numbers of symbionts observed in the parents. It would also be interesting to quantify the tissue-dependent distribution of the Δψ*adhE* symbionts compared to the WT symbionts at various stages after microinjection. This might help elucidate whether there are tissue-dependent effects, either due to colonization abilities of strains introduced via different routes or host responses to symbiont presence in different tissues.

Another possible explanation for loss of symbionts is that the inoculation procedure introduced an agent, a bacteriophage for example, that may have killed symbionts in some insects. However, this seems unlikely, given the manner in which symbiont loss occurred, with higher numbers of aposymbiotic insects observed in the F2 generation than the F1. Furthermore, this does not account for the preferential loss of WT symbionts, given that the two symbiont strains are not expected to differ in terms of phage sensitivity.

The results presented in the current study show that it is possible to replace native symbionts with genetically modified symbionts, without prior use of antibiotics and with no detrimental effects on the primary symbiont. Such a technique holds great promise for furthering our understanding of the molecular interactions between symbionts and their insect hosts, by facilitating the investigation of gene function within a natural insect host over the course of multiple generations. Furthermore, this technique holds promise for the development of paratransgenic techniques for the control of insect disease vectors such as tsetse flies. Obtaining flies lacking *Arsenophonus* symbionts altogether is also a useful outcome, providing a means by which to examine the role these bacteria play in their symbiotic association and how their presence affects louse fly survival and reproduction under different environmental conditions.

Acknowledgments

This manuscript will be submitted for publication with Judit Barabas and Colin Dale as coauthors. We thank Kelly Hughes for providing plasmid pKD46-CM and Dale Clayton for provision of louse flies. All procedures followed guidelines of the Institutional Animal Care and Use Committee of the University of Utah.

References

Aksoy, S. 1995. Wigglesworthia gen. nov. and Wigglesworthia glossinidia sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. *Int J Syst Bacteriol*, 45:848-51.

Aksoy, S., I. Maudlin, C. Dale, A. S. Robinson, and S. L. O'Neill. 2001. Prospects for control of African trypanosomiasis by tsetse vector manipulation. *Trends Parasitol*, 17:29-35.

Aksoy, S., B. Weiss, and G. Attardo. 2008. Paratransgenesis applied for control of tsetse transmitted sleeping sickness. *Adv Exp Med Biol*, 627:35-48.

Anselme, C., V. Perez-Brocal, A. Vallier, C. Vincent-Monegat, D. Charif, A. Latorre, A. Moya, and A. Heddi. 2008. Identification of the weevil immune genes and their expression in the bacteriome tissue. *BMC Biol*, 6:43.

Attardo, G. M., C. Lohs, A. Heddi, U. H. Alam, S. Yildirim, and S. Aksoy. 2008. Analysis of milk gland structure and function in Glossina morsitans: milk protein production, symbiont populations and fecundity. *J Insect Physiol*, 54:1236-42.

Beard, C. B., R. V. Durvasula, and F. F. Richards. 1998. Bacterial symbiosis in arthropods and the control of disease transmission. *Emerg Infect Dis*, 4:581-91.

Beard, C. B., S.L. O'Neill, P. Mason, L. Mandelco, C. R. Woese, R. B. Tesh, F. F. Richards, and S. Aksoy. 1993. Genetic transformation and phylogeny of bacterial symbionts from tsetse. *Insect Mol Biol*, 1:123-31.

Belda, E., A. Moya, S. Bentley, and F. J. Silva. 2010. Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of Sodalis glossinidius, the secondary endosymbiont of tsetse flies. *BMC Genomics*, 11:449.

Bequaert, J. The Hippoboscidae or louse-flies (Diptera) of mammals and birds. Part I. Structure, physiology and natural history. 1953. *Entomol Am* 32:1-209.

Blattner, F. R., G. Plunkett, 3rd, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of Escherichia coli K-12. *Science*, 277:1453-62.

Buchner, P. 1965. *Endosymbiosis of Animals with Plant Microorganisms*. John Wiley, New York.

Cheng, Q., and S. Aksoy. 1999. Tissue tropism, transmission and expression of foreign

genes in vivo in midgut symbionts of tsetse flies. *Insect Mol Biol*, 8:125-32.

Dale, C., M. Beeton, C. Harbison, T. Jones, and M. Pontes. 2006. Isolation, pure culture, and characterization of "Candidatus Arsenophonus arthropodicus," an intracellular secondary endosymbiont from the hippoboscid louse fly Pseudolynchia canariensis. *Appl Environ Microbiol*, 72:2997-3004.

Dale, C., and I. Maudlin. 1999. Sodalis gen. nov. and Sodalis glossinidius sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly Glossina morsitans morsitans. *Int J Syst Bacteriol*, 49 Pt 1:267-75.

Dale, C., and N. A. Moran. 2006. Molecular interactions between bacterial symbionts and their hosts. *Cell*, 126:453-65.

Dale, C., and S. C. Welburn. 2001. The endosymbionts of tsetse flies: manipulating host-parasite interactions. *Int J Parasitol*, 31:628-31.

Darby, A. C., S. M. Chandler, S. C. Welburn, and A. E. Douglas. 2005. Aphid-symbiotic bacteria cultured in insect cell lines. *Appl Environ Microbiol*, 71:4833-9.

Datsenko, K. A., and B. L. Wanner. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci USA*, 97:6640-5.

De Gelder, L., J. M. Ponciano, P. Joyce, and E. M. Top. 2007. Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. *Microbiology*, 153:452-63.

Derbise A., B. Lesic, D. Dacheux, J. M. Ghigo, and E. Carniel. 2003. A rapid and simple method for inactivating chromosomal genes in Yersinia. *FEMS Immunol Med Microbiol,* 38:113-116.

Douglas, A. E. 1998. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera. *Annu Rev Entomol*, 43:17-37.

Durvasula, R. V., A. Gumbs, A. Panackal, O. Kruglov, J. Taneja, A. S. Kang, C. Cordon-Rosales, F. F. Richards, R. G. Whitham, and C. B. Beard. 1999. Expression of a functional antibody fragment in the gut of Rhodnius prolixus via transgenic bacterial symbiont Rhodococcus rhodnii. *Med Vet Entomol*, 13:115-9.

Feldhaar, H. 2011. Bacterial symbionts as mediators of ecologically important traits of insect hosts. *Ecol Entomol*, 36:533-543.

Gherna, R. L., J. H. Werren, W. Weisburg, R. Cote, C. R. Woese, L. Mandelco, and D.J. Brenner. 1991. Arsenophonus nasoniae gen. nov., sp. nov., the causative agent of the son-killer trait in the parasitic wasp Nasonia vitripennis. *Int J Syst Bacteriol*,

41:563-565.

Hosokawa, T., Y. Kikuchi, and T. Fukatsu. 2007. How many symbionts are provided by mothers, acquired by offspring, and needed for successful vertical transmission in an obligate insect-bacterium mutualism? *Mol Ecol*, 16:5316-25.

Hypsa, V., and C. Dale. 1997. In vitro culture and phylogenetic analysis of "*Candidatus* Arsenophonus triatominarum," and intracellular bacterium from the triatomine bug, *Triatoma infestans*. *Int J Syst Bacteriol*, 47:1140-1144.

Leak, S. G. A. 1999. Tsetse biology and ecology; their role in the epidemiology and control of trypanosomiasis. CABI Publishing, Wallingford.

Login, F. H., S. Balmand, A. Vallier, C. Vincent-Monegat, A. Vigneron, M. Weiss-Gayet, D. Rochat, and A. Heddi. 2011. Antimicrobial peptides keep insect endosymbionts under control. *Science*, 334:362-5.

McCutcheon, J. P. 2010. The bacterial essence of tiny symbiont genomes. *Curr Opin Microbiol*, 13:73-8.

McCutcheon, J. P. and N. A. Moran. 2012. Extreme genome reduction in symbiotic bacteria. *Nature Rev Microbiol,* 10:13-26.

Mira, A., and N. A. Moran. 2002. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb Ecol*, 44:137-43.

Moran, N. A. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A*, 93:2873-8.

Moran, N. A., J. P. McCutcheon, and A. Nakabachi. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet*, 42:165-90.

Moran, N. A., and G.R. Plague. 2004. Genomic changes following host restriction in bacteria. *Curr Opin Gen Dev*, 14:627-633.

Nogge, G. 1981. Significance of symbionts for the maintenance of an optional nutritional state for successful reproduction in hematophagous arthropods. *Parasitol*, 82:101-104.

Novakova, E., V. Hypsa, and N. A. Moran. 2009. Arsenophonus, an emerging clade of intracellular symbionts with a broad host distribution. *BMC Microbiol*, 9:143.

Pais, R., C. Lohs, Y. Wu, J. Wang, and S. Aksoy. 2008. The obligate mutualist Wigglesworthia glossinidia influences reproduction, digestion, and immunity processes of its host, the tsetse fly. *Appl Environ Microbiol*, 74:5965-74.

Pontes, M. H., and C. Dale. 2011. Lambda Red-mediated genetic modification of the insect endosymbiont Sodalis glossinidius. *Appl Environ Microbiol*, 77:1918-20.

Pontes, M. H., K. L. Smith, L. De Vooght, J. Van Den Abbeele, and C. Dale. 2011. Attenuation of the sensing capabilities of PhoQ in transition to obligate insect-bacterial association. *PLoS Genet*, 7:e1002349.

Purcell, A. H., and R. P. P. Almeida. 2005. Insects as vectors of disease agents, *Encyclopedia of Plant and Crop Science*. Taylor & Francis, New York.

Reynolds, S., and J. Rolff. 2008. Immune function keeps endosymbionts under control. *J Biol*, 7:28.

Rio, R. V., Y. N. Wu, G. Filardo, and S. Aksoy. 2006. Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora. *Proc Biol Sci*, 273:805-14.

Sabri, A., P. Leroy, E. Haubruge, T. Hance, I. Frere, J. Destain, and P. Thonart. 2010. Isolation, pure culture and characterization of Serratia symbiotica sp. nov., the R-type of secondary endosymbiont of the black bean aphid Aphis fabae. *Int J Syst Evol Microbiol*, 61:2081-8.

Shigenobu, S., H. Watanabe, M. Hattori, Y. Sakaki, and H. Ishikawa. 2000. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. *Nature*, 407:81-6.

Weiss, B. L., R. Mouchotte, R. V. Rio, Y. N. Wu, Z. Wu, A. Heddi, and S. Aksoy. 2006. Interspecific transfer of bacterial endosymbionts between tsetse fly species: infection establishment and effect on host fitness. *Appl Environ Microbiol*, 72:7013-21.

CHAPTER 5

CONCLUSION

The work presented here describes the development of a new tractable system for investigating mutualistic insect-bacterial symbioses, which are widespread across many insect lineages. Due to the high levels of specialization common to partners in these types of associations, there are few bacterial symbionts that have yet proved amenable to culture and genetic manipulation in the lab. Thus, much of the study of insect symbionts has been genomics-based, and while there is much genomic information available for symbionts, the development of genetic techniques to investigate gene function in these interactions is still lagging. There is a definite need to develop new platforms to begin to understand the molecular mechanisms underlying the initiation and maintenance of long-term mutualistic insect-bacterial associations.

The completion of the genome sequence of *Candidatus* Arsenophonus arthropodicus provides a valuable addition in that it presents a snapshot of the genomic degeneration that occurs after bacteria become associated with an insect host, and appears to represent a symbiont that has become associated with insects only recently. Furthermore, the availability of a genome sequence for a very close relative, the reproductive parasite *Arsenophonus nasoniae* (Darby et al., 2010; Wilkes et al., 2010), that has a lifestyle quite distinct from that of *Ca.* A. arthropodicus, provides an interesting complement for comparative studies of genome evolution and insect interactions.

Comparative investigation of these two symbionts may provide some insight into the means by which members of the *Arsenophonus* genus can infect such a wide range of insects with diverse lifestyles and host effects (Wilkes et al., 2011).

The generation of genome sequences for these symbionts has also provided a variety of interesting gene targets, such as genes involved in type III secretion, flagellar motility, or toxin compounds, which need to be experimentally tested to determine their necessity within the symbiosis of *Ca.* A. arthropodicus with its louse fly host. The successful application of lambda Red recombineering methods (Datsenko and Wanner, 2000) to *Ca.* A. arthropodicus allows us now to knock out specific genes within the bacterium and examine the effects both *in vitro* and *in vivo*. Furthermore, the generation of a phenotypically neutral strain (Δψ*adhE)*, enables a novel method of testing and quantifiying mutant infection ability *in vivo*, by competing it against new mutant strains to determine the ability of the mutant strain to initiate and maintain an association as compared to the phenotypically wild-type Δψ*adhE* strain. The generation of targeted knockout strains of *Ca.* A. athropodicus is currently underway, with a recombinant symbiont strain lacking an essential flagellar gene, *fliM*, having undergone microinjection into louse flies in competition with the Δψ*adhE* strain to determine if flagella production is important for maintenance of the symbiosis. Other genes should also be tested in this manner to determine their utility in the association.

This work also describes the replacement of native symbionts in an insect host with genetically modified symbionts via microinjection, without the use of antibiotic or other treatments to first reduce native symbiont populations. This enables the establishment of a recombinant symbiont strain without affecting the obligate symbionts

necessary for louse fly survival and reproduction, providing a unique method for introducing and examining recombinant symbionts *in vivo*. In addition, a side effect of the microinjection procedure was the procurement, for the first time, of louse flies that were aposymbiotic for *Ca.* A. arthropodicus. This result now provides the opportunity to investigate the role of the symbiont within its host, by comparing the fitness of aposymbiotic versus symbiont-containing flies under various environmental conditions, nutrient availabilities or parasitic pressures.

The successful infection of symbionts after microinjection into louse flies also provides a means by which to investigate the abilities of various symbionts to infect and establish in novel hosts and determine the capability of symbionts to undergo horizontal transfer between different insect species. Symbionts isolated from other insects can be microinjected into louse fly pupae to determine if they can initiate and maintain an infection, and the competitive ability of novel symbionts can be compared to native symbionts within louse flies. Future work could entail investigating both the independent and competitive abilities of *A. nasoniae* and *Ca.* A. arthropodicus to sustain infections in louse fly and parasitoid wasp hosts and identify the specific effects each species has on its own host vs. a novel host.

The development of methods to generate recombinant symbionts and establish them in an insect host also provides a means to further develop techniques for paratransgenesis. This is an important avenue of emerging research, given that many insects harbor and transmit parasites that cause disease in humans or other animals (Coutinho-Abreu et al., 2010). There is much interest in controlling parasites through insect transgenesis; however, genetically modifying insects to express genes that inhibit

parasite transmission is a very difficult feat and taking advantage of insect symbionts to express transgenes might present a more feasible route of experimentation (Coutinho-Abreu et al., 2010). Advancement on this front has been hampered by the lack of appropriate study systems consisting of culturable symbionts, techniques for generating stable genetic modifications in the symbionts and establishing recombinant symbionts in insects hosts. Thus, this system provides a new platform with which to pursue these goals as well as to begin to unravel the molecular basis underlying the evolution of mutualistic symbioses between bacteria and insects.

References

Coutinho-Abreu, I. V., K. Y. Zhu, and M. Ramalho-Ortigao. 2010. Transgenesis and paratransgenesis to control insect-borne diseases: current status and future challenges. *Parasitol Int*, 59:1-8.

Darby, A. C., J. H. Choi, T. Wilkes, M. A. Hughes, J. H. Werren, G. D. Hurst, and J. K. Colbourne. 2010. Characteristics of the genome of Arsenophonus nasoniae, son-killer bacterium of the wasp Nasonia. *Insect Mol Biol*, 19 Suppl 1:75-89.

Datsenko, K. A., and B. L. Wanner. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A*, 97:6640-5.

Wilkes, T. E., A. C. Darby, J. H. Choi, J. K. Colbourne, J. H. Werren, and G. D. Hurst. 2010. The draft genome sequence of Arsenophonus nasoniae, son-killer bacterium of Nasonia vitripennis, reveals genes associated with virulence and symbiosis. *Insect Mol Biol*, 19 Suppl 1:59-73.

Wilkes, T. E., O. Duron, A. C. Darby, V. Hypsa, E. Novakova, and G. D. D. Hurst. 2011. The genus Arsenophonus. In E. Zchori-Fein, and K. Bourtzis (ed.), *Manipulative tenants: Bacteria associated with arthropods*. CRC Press, Boca Raton, FL.

APPENDIX A

ATTENUATION OF THE SENSING CAPABILITIES OF PHOQ IN TRANSITION TO OBLIGATE INSECT-BACTERIAL ASSOCIATION

Reprinted with permission from:
Pontes, M. H., K. L. Smith, L. De Vooght, J. Van Den Abbeele, and C. Dale. 2011. Attenuation of the sensing capabilities of PhoQ in transition to obligate insect-bacterial association. *PLoS Genet*, 7:e1002349.

OPEN ACCESS Freely available online

PLoS GENETICS

Attenuation of the Sensing Capabilities of PhoQ in Transition to Obligate Insect–Bacterial Association

Mauricio Henriques Pontes[1¤], Kari Lyn Smith[1], Linda De Vooght[2], Jan Van Den Abbeele[2], Colin Dale[1]*

1 Department of Biology, University of Utah, Salt Lake City, Utah, United States of America, **2** Department of Biological Sciences, Institute of Tropical Medicine, Antwerp, Belgium

Abstract

Sodalis glossinidius, a maternally inherited endosymbiont of the tsetse fly, maintains genes encoding homologues of the PhoP-PhoQ two-component regulatory system. This two-component system has been extensively studied in facultative bacterial pathogens and is known to serve as an environmental magnesium sensor and a regulator of key virulence determinants. In the current study, we show that the inactivation of the response regulator, *phoP*, renders *S. glossinidius* sensitive to insect derived cationic antimicrobial peptides (AMPs). The resulting mutant strain displays reduced expression of genes involved in the structural modification of lipid A that facilitates resistance to AMPs. In addition, the inactivation of *phoP* alters the expression of type-III secretion system (TTSS) genes encoded within three distinct chromosomal regions, indicating that PhoP-PhoQ also serves as a master regulator of TTSS gene expression. In the absence of *phoP*, *S. glossinidius* is unable to superinfect either its natural tsetse fly host or a closely related hippoboscid louse fly. Furthermore, we show that the *S. glossinidius* PhoQ sensor kinase has undergone functional adaptations that result in a substantially diminished ability to sense ancestral signals. The loss of PhoQ's sensory capability is predicted to represent a novel adaptation to the static symbiotic lifestyle, allowing *S. glossinidius* to constitutively express genes that facilitate resistance to host derived AMPs.

Citation: Pontes MH, Smith KL, De Vooght L, Van Den Abbeele J, Dale C (2011) Attenuation of the Sensing Capabilities of PhoQ in Transition to Obligate Insect–Bacterial Association. PLoS Genet 7(11): e1002349. doi:10.1371/journal.pgen.1002349

Editor: Nancy A. Moran, Yale University, United States of America

Received July 7, 2011; **Accepted** August 30, 2011; **Published** November 3, 2011

Funding: This research was supported by National Science Foundation (www.nsf.gov) Grant EF-0523818 (to CD). MHP was supported by a Stringfellow Award at the University of Utah. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: colin.dale@utah.edu

¤ Current address: Section of Microbial Pathogenesis, Howard Hughes Medical Institute, Yale University School of Medicine, Boyer Center for Molecular Medicine, New Haven, Connecticut, United States of America

Introduction

Many animals have adopted mutualistic associations with bacteria. These associations are based on an exchange in which the bacterial symbiont provides functions that enhance host survival, and the host provides a stable, nutrient-rich home for the bacterial symbiont. Over the course of macroevolutionary time, the metabolic and physiological activities of the host and symbiont become increasingly integrated, leading to an obligate mutualism. Under these conditions, the host cannot survive without the ancillary functions provided by the bacterial symbiont and the bacterium cannot persist outside of the host. The reliance of the bacterial symbiont on the animal host is elegantly illustrated in a number of insect-bacterial symbioses. In these associations, dependency arises as a result of bacterial genome degeneration and size reduction [1]. In extreme cases, when the symbiotic partners have co-evolved for long periods of time, bacterial endosymbionts display pronounced genome streamlining. For example, bacterial endosymbionts of aphids, *Buchnera* spp., have genome sizes ranging from 0.42 to 0.66 Mbp, that are substantially smaller than those of close, free-living relatives (e.g. *Escherichia coli* K-12; 4.6 Mbp) [2].

Although genome streamlining is most conspicuous in ancient associations, the early stages of this process have been observed in symbiotic associations that are more recent in origin. For example, the relationship between tsetse flies and their bacterial endosymbiont *Sodalis glossinidius* is predicted to be relatively recent in origin [3], and although the genome size of *S. glossinidius* (4.29 Mbp) is similar to that of close, free-living relatives (e.g. *Yersinia enterocolitica*; 4.6 Mbp) [4], a significant portion of the *S. glossinidius* genome is composed of pseudogenes that have been inactivated as a consequence of relaxed selection, because they no longer play a vital role in the symbiosis [5]. However, despite this extensive genome degeneration, several recently derived endosymbionts (including *S. glossinidius*) have been shown to maintain intact copies of genes sharing high levels of sequence identity with homologs encoding virulence determinants in common plant and animal pathogens [6–10]. Since these virulence genes are not found in ancient mutualistic endosymbionts, it has been suggested that they play a transient role in the establishment of these symbiotic associations [1]. Notably, like many plant and animal pathogens, recently derived insect endosymbionts also maintain an extensive repertoire of regulatory genes [5,11]. However, while pathogens use these regulators to rapidly coordinate adaptations to diverse environments (e.g. host vs. non-host), the mutualistic endosymbionts of insects are permanently host associated and are therefore entrenched in a more static lifestyle. The ecological changes associated with a lifestyle switch from opportunism to obligate host association are therefore predicted to mediate adaptive changes in the functions of regulatory circuits that serve as environmental sensors.

Many Gram-negative pathogens utilize the PhoP-PhoQ two-component regulatory system to modulate adaptive responses to changes in levels of divalent cations, including magnesium, in the

Author Summary

Mutualistic insect endosymbionts are known to undergo a process of degenerative evolution that streamlines their gene inventory in accordance with the obligate nature of the host associated lifestyle. Here we show that the mutualistic insect endosymbiont *Sodalis glossinidius* utilizes a two-component regulatory system (PhoP-PhoQ) that evolved to facilitate the regulation of virulence genes in facultative pathogens in accordance with environmental magnesium. While the regulatory targets of PhoP-PhoQ are conserved among *S. glossinidius* and a wide range of bacterial pathogens, the *S. glossinidius* PhoP-PhoQ system is unusual because it has a substantially reduced ability to sense magnesium as a repressing signal. This enables *S. glossinidius* to express genes encoding critical symbiosis determinants within host environments that are rich in both magnesium and antimicrobial peptides. These findings illustrate how a change in the ecology of a host–symbiont association can influence the dynamic functionality of a regulatory system in the symbiont.

environment [12 14]. When magnesium availability is high, the inner membrane sensor kinase PhoQ dephosphorylates the cytoplasmic response regulator PhoP, maintaining the system in a deactivated state. When magnesium availability is low, PhoQ autophosphorylates and transfers its phosphoryl group to PhoP. Phosphorylated PhoP then activates the expression of target genes that are associated with adaptation to the low magnesium environment [13,14]. In multicellular eukaryotes, intra and extracellular concentrations of magnesium vary considerably. While extracellular concentrations are in the millimolar range, intracellular concentrations tend to be in the micromolar range [12,13,15]. Notably, in many facultative intracellular pathogens, PhoP-PhoQ regulates the expression of genes important for intracellular survival. Low intracellular magnesium levels drive PhoP-PhoQ-dependent expression of loci involved in magnesium transport [13,14] and structural modifications of the lipid A portion of the bacterial lipopolysaccharide (LPS) [16]. Whereas magnesium transport genes allow bacteria to obtain adequate amounts of magnesium for survival, LPS modifications protect the bacteria against stressful conditions found within eukaryotic phagosomes, such as low pH and high levels of antimicrobial peptides (AMPs) [13,15]. Notably, in addition to low magnesium, the PhoP-PhoQ system is also known to detect and respond to other host derived signals found within phagosomes, such as AMPs [17] and acidic pH [18]. In contrast to magnesium, which inhibits PhoP-PhoQ, the binding of AMPs or the exposure of cells to acidic conditions results in the activation of PhoP-PhoQ [17,18].

At present, very little is known about how mutualistic endosymbionts evade or overcome the challenges imposed by the insect immune system. The immune systems of multicellular organisms utilize a vast array of mechanisms to combat invading microorganisms. The immune cells of both insects and vertebrates are known to synthesize various cationic AMPs that kill bacteria by interacting with lipid A and forming holes into the bacterial lipid membrane. In insects, these immune peptides combat bacterial pathogens by functioning as antibiotics that are secreted into the hemolymph and stored within phagocytic cells, where they are used to kill engulfed bacteria [19,20]. To date, only two insect endosymbionts (*S. glossinidius* and *Candidatus* Arsenophonus arthropodicus) have been cultured in the laboratory and tested for resistance to cationic AMPs; notably, both of these endosymbionts were found to display high levels of resistance *in vitro* [21 23].

In the current study, we show that the mutualistic insect endosymbiont *S. glossinidius* utilizes a PhoP-PhoQ two-component regulatory system to modulate the expression of genes involved in lipid A modifications that confer bacterial resistance to host derived AMPs. In the absence of PhoP, *S. glossinidius* demonstrates increased sensitivity to host derived AMPs, an aberrant profile of type-III secretion system (TTSS) gene expression and an inability to colonize its natural host, the tsetse fly, and a close dipteran relative, the hippoboscid louse fly. In addition, our results indicate that the PhoP-PhoQ system of *S. glossinidius* has undergone sensory adaptations in the transition to a permanent association with its insect host.

Results

PhoP Is Required for Resistance to Cationic AMPs *In Vitro*

In previous studies, *S. glossinidius* was shown to be highly resistant to the effects of a number of cationic AMPs [22,23]. In the current study we examined the sensitivity of a *S. glossinidius phoP* mutant [24] to polymyxin B and the insect immune peptide, cecropin A, which is known to be produced by the host of *S. glossinidius* [25,26]. Whereas wild type *S. glossinidius* demonstrated high levels of resistance as expected, the *phoP* mutant strain was found to be extremely sensitive (c. 1000-fold increase in sensitivity) to both AMPs (Figure 1). Resistance to polymyxin B, but not cecropin A, was slightly increased in response to low magnesium availability in the *S. glossinidius* culture medium (Figure 1). However, this effect was mediated in both the wild type and *phoP* mutant strains, indicating that it is PhoP-independent. These results show that PhoP-PhoQ plays a vital role in mediating AMP resistance in *S. glossinidius*.

PhoP Activates Expression of a Gene Known to Mediate Resistance to Cationic AMPs

In facultative intracellular pathogens, resistance to AMPs is mediated by structural modifications of the lipid A portion of the bacterial LPS [12,13]. Because cationic AMPs kill bacteria by interacting with unmodified lipid A and interfering with the permeability of the bacterial lipid membrane [17,19], we reasoned that the *S. glossinidius phoP* mutant might lack the ability to carry out the necessary structural modifications of lipid A. Inspection of the *S. glossinidius* genome sequence revealed the presence of three loci (*pagP*, locus tag SG1577; *pmrE*, locus tag SG1368; and *pmrH*, the first gene in the *pmrHFIJKLM* operon, locus tags SG1845 to SG1839, respectively) encoding proteins known to mediate modifications that reduce the negative charge of lipid A, preventing the binding of positively charged AMPs [13]. PagP mediates the palmytoylation of lipid A, a structural modification associated with bacterial resistance to alpha-helical AMPs such as cecropin A. PmrE and the proteins encoded by the *pmrH* operon mediate the synthesis and incorporation of 4-aminoarabinose into the lipid A, a modification associated with bacterial resistance to the cyclic lipopeptide polymyxin B [27].

Our quantitative PCR results show that *pagP* and *pmrE* are not regulated by PhoP (Figure 2). Instead, both genes were found to display high levels of expression under all conditions tested (data not shown), suggesting that they have likely evolved to a state of constitutive expression. However, the expression of the *pmrH* operon was found to be PhoP-dependent in *S. glossinidius* (Figure 2). When cells were grown in medium with high levels of magnesium, wild type *S. glossinidius* showed 29-fold higher levels of *pmrH* mRNA relative to the *phoP* mutant (Figure 2). Furthermore, when cells were grown in medium containing low levels of magnesium, the levels of *pmrH* mRNA were 63-fold higher in wild type *S.*

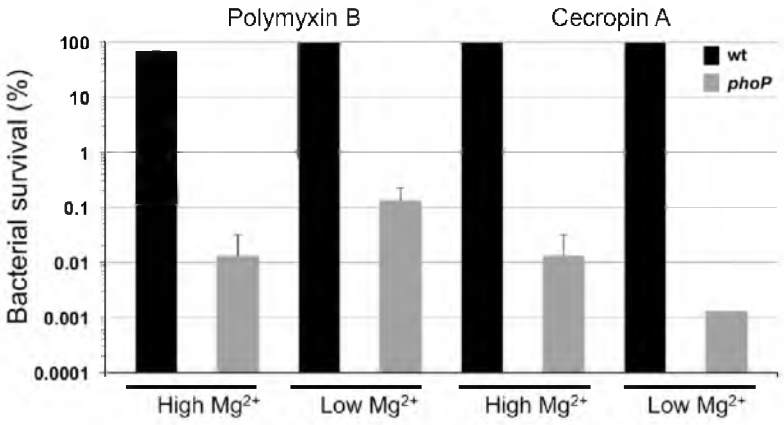


Figure 1. Resistance to polymyxin B and cecropin A is PhoP-dependent in *S. glossinidius*. Whereas magnesium has no effect on *S. glossinidius* wild type (wt) resistance to cecropin A, it has a small but significant, PhoP-independent effect on resistance to polymyxin B (two-tail t-test: $p<0.01$). Wild type and *phoP* mutant cells were cultured at high (10 mM) and low (10 μM) levels of magnesium, exposed to antimicrobial peptides and enumerated after plating. Error bars represent standard deviations.
doi:10.1371/journal.pgen.1002349.g001

glossinidius relative to the *phoP* mutant (Figure 2). Together, these results indicate that *S. glossinidius* uses the PhoP-PhoQ system to activate the transcription of the *pmrH* operon and direct modifications of lipid A that are important determinants of resistance to both polymyxin B and cecropin A.

Sodalis glossinidius phoP Mutant Displays an Altered Lipid Composition

Because PhoP-PhoQ often govern the expression of genes involved in the structural modifications of the outer surface of the bacterial lipid membrane [16,28] we sought to determine if the genetic inactivation of *phoP* resulted in changes in the overall lipid composition of *S. glossinidius* cells. We carried out a thin layer chromatography (TLC) analysis of lipids derived from the wild type and *phoP* mutant strains of *S. glossinidius*. Our TLC analysis shows that the wild type *S. glossinidius* produces an additional (as yet uncharacterized) lipid species that is absent in the *phoP* mutant strain (Figure 3). This result reinforces the notion that *S. glossinidius* also utilizes the PhoP-PhoQ system to regulate expression of genes involved in lipid metabolism.

PhoP Also Regulates Type-III Secretion System Genes in *Sodalis glossinidius*

Some PhoP regulated promoters have a characteristic direct repeat sequence, known as a PhoP box, which serves as a binding site for phosphorylated PhoP [29,30]. Inspection of the *S. glossinidius* genome sequence revealed the presence of a consensus PhoP box upstream of *hilA* (accession no. AAS66857; Figure 4), which is known to encode a master regulator of type-III secretion in *Salmonella enterica* [31]. Because the *S. glossinidius* genome has three distinct symbiosis regions (designated *Sodalis* Symbiosis Regions; SSR's) encoding TTSS genes [5], we elected to measure the effect of *phoP* inactivation on the basal expression levels of the genes found in these three distinct chromosomal locations using quantitative PCR. The *phoP* mutant was found to have significantly lower levels of transcripts encoding YsaE, SycB (SSR-1), HilA, InvF, SicA (SSR-2), SsrB, SsaB, SseB and SsaH (SSR-3; Figure 2). These results indicate that the *S. glossinidius* PhoP-PhoQ two-component system is also involved in the activation of TTSS genes that are known to be required for the invasion of insect cells and for intracellular survival [8].

Sodalis glossinidius phoP Mutant Fails to Superinfect Insect Hosts

One important component of insect immunity involves the synthesis and secretion of high quantities of AMPs into the hemolymph [20]. Since insect endosymbionts, including *S. glossinidius*, are often found in the hemolymph of their insect hosts [32,33], and because *S. glossinidius* uses TTSSs to invade and replicate within insect cells [6,8], we assessed the ability of the *phoP* mutant strain of *S. glossinidius* to superinfect their natural tsetse fly host, *Glossina morsitans morsitans*, and a closely-related hippoboscid louse fly, *Pseudolynchia canariensis*, following intrathoracic microinjection. The superinfection approach [34] was selected to avoid the requirement to treat host insects with antibiotics to remove native symbionts, because this procedure is known to compromise the immune system of the fly [35]. The presence of native *S. glossinidius* in the tsetse fly therefore mandated the use of a recombinant (*fliM* mutant) *S. glossinidius* strain in lieu of a wild type control strain so that microinjected bacteria could be differentiated from native symbionts by PCR. In comparison to the *fliM* mutant strain, the *phoP* mutant strain was found to be completely defective in its ability to superinfect the host insect (Table 1). At seven days following microinjection, we were able to detect the *fliM* mutant strain in 26 out of 32 tsetse flies sampled. In contrast, we were unable to detect either the *phoP* or *fliM* heat-killed mutant strains in any of the tsetse flies sampled at seven days post-microinjection (Table 1). In the hippoboscid louse flies, which are closely related to tsetse flies [36] but do not maintain a native population of *S. glossinidius*, we compared the superinfection abilities of wild type, *fliM* and *phoP* mutant strains of *S. glossinidius*. In this experiment, wild type and *fliM* mutant *S. glossinidius* were detected in all flies sampled at seven days following microinjection, and the *phoP* mutant and heat-killed strains of *S. glossinidius* were not detected in any flies at seven days post injection (Table 1). Based on these results, we conclude that the *phoP* mutant strain of *S. glossinidius* is killed by the insect immune system following superinfection in both tsetse flies and hippoboscid louse flies. This indicates that the PhoP-PhoQ two-component regulatory system is essential for the

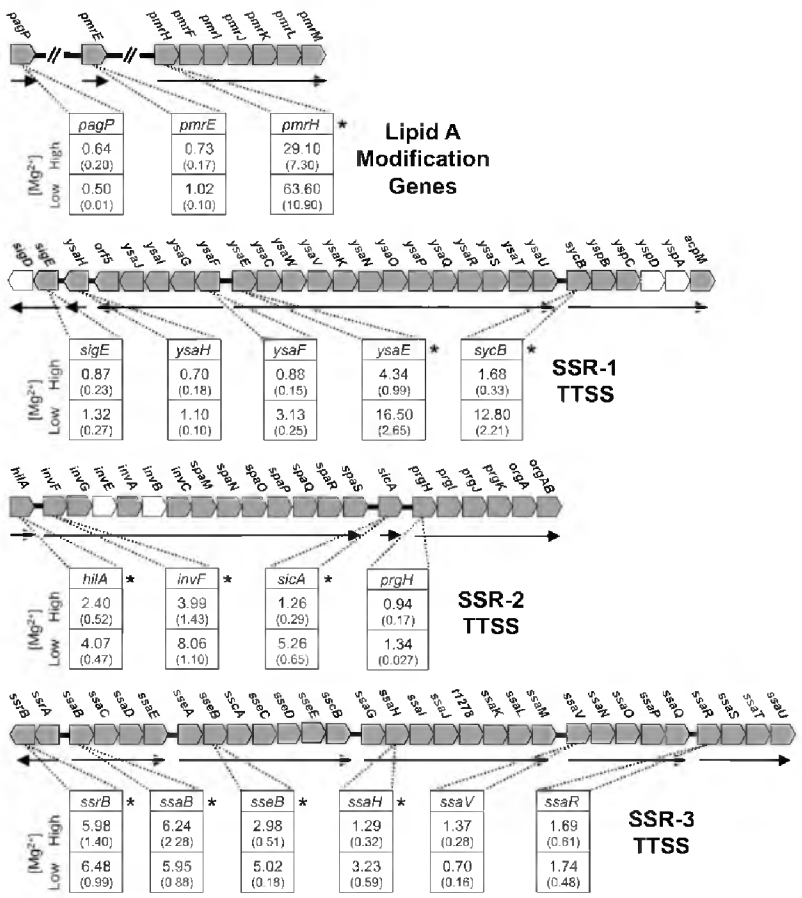


Figure 2. Quantitative PCR analysis of transcripts derived from genes involved in lipid A modifications in *S. glossinidius*. Genetic organization of lipid A modification genes is depicted: Functional genes are decorated in gray and putative transcriptional units are indicated by arrows. The numbers in boxes indicate the ratios of transcripts detected in the wild type strain relative to the *phoP* mutant strain of *S. glossinidius* grown under conditions of high (10 mM) and low (10 μM) magnesium availability. Genes displaying significant changes in expression in the wild type relative to the *phoP* mutant strain of *S. glossinidius* are highlighted with asterisks (two-tail t-test: $p<0.05$). Values in parentheses represent standard deviations.
doi:10.1371/journal.pgen.1002349.g002

establishment and maintenance of an *S. glossinidius* infection in an insect host.

Effects of Acidic pH, Magnesium, and Cationic AMPs on *Sodalis glossinidius*

In many facultative bacterial pathogens, the PhoP-PhoQ system functions as a magnesium sensor that controls the expression of genes mediating physiological adaptations to changes in environmental levels of magnesium [13-15]. In addition, the PhoQ sensor kinase is known to detect and respond to conditions of acidic pH [18], and the binding of cationic AMPs that displace magnesium ions [17]. Several lines of evidence indicate that the *S. glossinidius* PhoP-PhoQ system has a diminished ability to respond to environmental magnesium. First, in *S. enterica* AMP resistance is controlled by the PhoP-PhoQ two-component system. *Salmonella enterica* cells grown under conditions of high magnesium availability are >1000-fold more susceptible to AMPs [37]. In contrast, our AMP resistance assay showed that although PhoP-PhoQ controls AMP resistance in *S. glossinidius*, this phenotype is only marginally affected by magnesium availability in the culture medium (Figure 1). Second, our gene expression analyses show that PhoP activates the transcription of *pmrH* and a number of TTSS genes in *S. glossinidius*. However, while the expression of some of these genes is considerably higher in wild type *S. glossinidius* relative to the *phoP* mutant strain, transcriptional changes in response to environmental magnesium were unexpectedly small. Finally, the results from our TLC lipid analysis showed that, regardless of magnesium concentration, wild type *S. glossinidius* synthesizes a lipid species that is not found in the *phoP* mutant strain (Figure 3).

To further explore the role of magnesium, acidic pH and the presence of AMPs on signaling mediated by the *S. glossinidius* PhoQ, we first used quantitative PCR to measure changes in the numbers of transcripts of *pmrH* in wild type cells under different culture conditions. The results show that *pmrH* expression was only slightly increased under conditions of low magnesium availability or low pH (2-fold and 2.6-fold increases, respectively). Further-

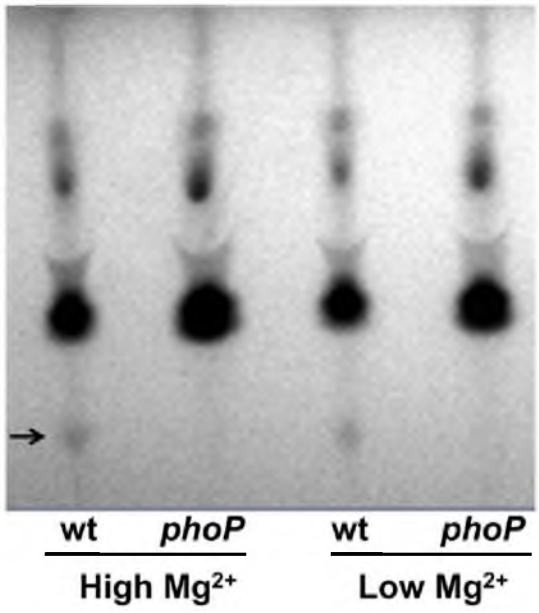


Figure 3. Thin layer chromatographic analysis of lipids extracted from wild type (wt) and *phoP* mutant strains of *S. glossinidius* grown at high (10 mM) and low (10 μM) concentrations of magnesium. Arrow highlights a lipid species that is absent in the *phoP* mutant strain.
doi:10.1371/journal.pgen.1002349.g003

more, the presence of the AMPs cecropin A or polymyxin B in the culture medium had no significant effect on *pmrH* gene expression (Figure 5A).

Since cellular adaptations mediated by PhoP-PhoQ are anticipated to be functionally linked to the physiological signals sensed by PhoQ, we also tested the abilities of wild type and *phoP* mutant strains of *S. glossinidius* to resist acid stress, and to grow under conditions of magnesium starvation. Notably, both wild type and *phoP* mutant strains of *S. glossinidius* showed the same level of sensitivity towards acid stress (Figure 5B). Similarly, no significant difference was observed in the abilities of wild type and *phoP* mutant strains of *S. glossinidius* to grow in a defined medium in which bacterial growth was clearly limited by a lack of magnesium (Figure 5C). Together, these results indicate that (i) the *S. glossinidius* PhoP-PhoQ system displays an atypical, diminished response to magnesium, acidic pH and AMPs, and (ii) the *S. glossinidius* PhoP response regulator does not facilitate tolerance towards acid stress or magnesium starvation in *S. glossinidius*, as it does in several enteric pathogens.

Salmonella enterica Strains Expressing *Sodalis glossinidius* PhoQ Are Magnesium-Insensitive

To further analyze the magnesium sensing capability of the *S. glossinidius* PhoQ sensor kinase we took advantage of an approach that was initially devised to study the magnesium binding property of the *S. enterica* PhoQ homologue [38]. We utilized two *S. enterica* strains (EG5931 and EG9461) both of which encode a *phoQ* null allele and a transcriptional fusion of the lac gene to a chromosomal copy of the *S. enterica* PhoP activated *pmrC* gene (*pmrC*::MudJ). Whereas one of these strains (EG9461) has a wild type *phoP* allele and can be used to measure induction levels of PhoP-activated genes at low magnesium availability, the other strain (EG5931) has a *phoP** allele encoding a protein that efficiently autophosphorylates from acetyl phosphate (i.e., independently from PhoQ). Because PhoP* responds normally to the phosphatase activity of PhoQ, strain EG5931 can be used to examine the capability of PhoQ to dephosphorylate PhoP and repress transcription of PhoP-activated genes [38,39].

We used strain EG9461 to test the ability of the *S. glossinidius* PhoQ sensor kinase to induce the expression of PhoP-activated genes in response to low magnesium availability. For controls, we complemented strain EG9461 with either a plasmid vector alone, a plasmid expressing the *S. enterica* wild type PhoQ sensor kinase (p*phoQ*) or a plasmid expressing a *S. enterica* PhoQ variant with mutation D150A (p*phoQ* D150A) in the magnesium ligand-binding site [38] (Figure S1). Strain EG9461 harboring the plasmid vector alone does not express a PhoQ protein and, therefore, produces no β-galactosidase activity from the *pmrC*::MudJ reporter fusion (Figure 6A). On the other hand, when this strain is complemented with a plasmid expressing the *S. enterica* wild type PhoQ sensor kinase (p*phoQ*), the expression of the reporter fusion mimics that of wild type cells; this strain displays high expression at low magnesium levels and low expression at high magnesium levels

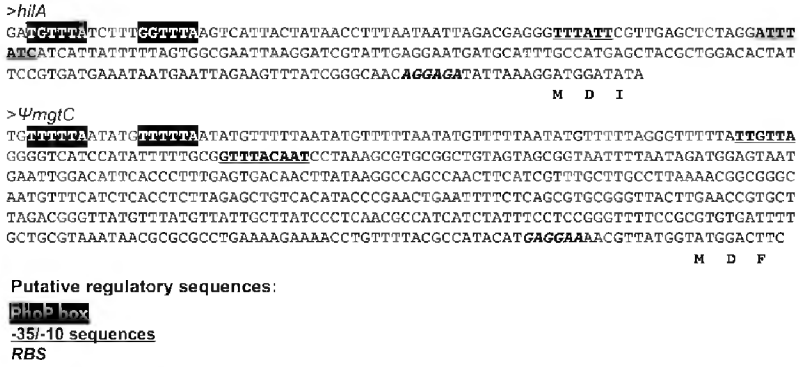


Figure 4. The putative promoter regions of the *S. glossinidius hilA* homologue and the *mgtCB* pseudo-operon contain canonical PhoP boxes. PhoP boxes (inverted text) and putative ribosomal binding site (bold) were identified by visual inspection of the promoter regions. Putative -35 and -10 regions (underlined) were identified using the online BPROM tool (SoftBerry, Inc.).
doi:10.1371/journal.pgen.1002349.g004

Table 1. PCR detection of *Sodalis glossinidius* seven days following microinjection in tsetse flies and louse flies.

S. glossinidius strain	Number of PCR positive samples/number of total samples	
	Tsetse flies	Louse flies
Wild-type	N.D.	19/20
Wild-type heat-killed	N.D.	0/20
fliM	26/32	16/16
***fliM* heat-killed**	0/32	N.D.
phoP	0/32	0/20

Differences in colonization patterns between the wild-type, *fliM* and *phoP* strains are statistically significant (Pearson X^2 test, $p<0.0001$). "N.D." indicates that that a particular experiment was not performed.
doi:10.1371/journal.pgen.1002349.t001

(Figure 6A). When complemented with a plasmid expressing the *S. enterica* PhoQ variant with mutation D150A (p*phoQ* D150A), strain EG9461 is derepressed under conditions of high magnesium availability (Figure 6A) [38]. In addition, when strain EG9461 is complemented with a plasmid expressing the *S. glossinidius* PhoQ protein (p*phoQ Sg*) there is an increase in β-galactosidase activity of the reporter fusion relative to that of strain EG9461 harboring the plasmid vector alone (Figure 6A), indicating that *S. glossinidius* PhoQ does phosphorylate the *S. enterica* PhoP response regulator. However, consistent with our previous observation, the resulting strain shows only a slight change in reporter gene activity in response to magnesium (Figure 6A).

Similar results were obtained in an experiment involving *S. enterica* strain EG5931, which expresses an autophosphorylating *phoP** allele. In this case strain EG5931, harboring the plasmid vector alone, displayed high levels of β-galactosidase activity from the reporter fusion, regardless of magnesium availability (Figure 6B). Yet, because the *phoP** allele responds normally to the phosphatase activity of PhoQ, complementation of strain EG5931 with a plasmid expressing the *S. enterica* wild type PhoQ (p*phoQ*) produced a strain that displayed a gene expression profile mimicking that of wild type cells. As with strain EG9461, complementation of strain EG5931 with a plasmid expressing the *S. enterica* PhoQ variant with mutation D150A (p*phoQ* D150A) gave rise to a strain that had a reduced capability to promote magnesium mediated repression of the *pmrC*::MudJ reporter fusion (Figure 6B) [38]. In addition, when strain EG5931 was complemented with a plasmid expressing the *S. glossinidius* wild type PhoQ protein (p*phoQ Sg*), the resulting strain again demonstrated little or no response to magnesium. Furthermore, this strain demonstrated decreased β-galactosidase activity of the reporter fusion relative to that of strain EG5931 harboring the plasmid vector alone (Figure 6B), indicating that the *S. glossinidius* wild type PhoQ sensor kinase does maintain a phosphatase activity. Together, these results further support our observation that the *S. glossinidius* PhoQ has a substantially diminished ability to sense magnesium.

Evidence for Magnesium Sensing in an Ancestral Precursor of the *S. glossinidius* PhoQ

Given that the majority of PhoP-PhoQ homologues studied in bacteria are known to respond aggressively to changes in acidic pH, magnesium and cationic AMPs [13,14,17,18], it is striking that the *S. glossinidius* PhoQ sensor kinase displays a substantially reduced ability to respond to these signals. In many facultative pathogens, PhoP-PhoQ is known to play an important role in mediating magnesium homeostasis [13,14]. When PhoQ senses conditions of low magnesium availability, PhoP responds by increasing expression of specialized, ATP-driven magnesium transporters (designated MgtA and MgtB in *S. enterica*) [14,40,41]. While *S. glossinidius* maintains intact copies of the generalized magnesium transporters *corA* (locus tag SG2341) and *mgtE* (locus tags SG0628 and SG1738) that are not regulated by PhoP-PhoQ in other bacteria, it completely lacks *mgtA* and maintains only a disrupted copy of *mgtB* with multiple frameshifts (Table 2). However, the disrupted copy of *mgtB*, encoded within the *mgtCB* pseudo-polycistron maintains a canonical PhoP box (Figure 4), implying that in the recent evolutionary past *S. glossinidius* used PhoP-PhoQ to coordinate the expression of genes involved in magnesium transport. Since it would be unexpected to have a magnesium transport system controlled by a PhoP-PhoQ system that is unable to sense magnesium, we conclude that an ancestral precursor of the *S. glossinidius* PhoQ protein possessed an increased magnesium sensing capability. This implies that the reduction in *S. glossinidius* PhoQ's ability to sense magnesium is a derived characteristic.

Discussion

The transition in lifestyle from opportunism to obligate host association is often accompanied by the inactivation and loss of genes that are assumed to have played an important role in a facultative lifestyle but no longer provide an adaptive benefit in an obligate host-associated lifestyle [1]. While these degenerative changes are not expected to negatively impact the function of the symbiotic relationship, they are anticipated to increase host dependence as a result of niche specialization on the part of the symbiont. Furthermore, the transition from a dynamic lifestyle to a static, obligate host-associated lifestyle is expected to reduce the requirement for symbionts to engage in regulatory activities that normally enhance the ability of bacteria to survive in a changing environment.

In the current study we explored the role of the two-component regulatory system, PhoP-PhoQ, in the insect endosymbiont, *S. glossinidius*, which is an obligate, mutualistic associate of tsetse flies [42]. PhoP-PhoQ is of particular interest because it is found in a wide range of facultative pathogens [13,14,16] and plays an important role in enabling these bacteria to sense their presence in the host environment and mediate changes in gene expression that facilitate important adaptations to the host-associated lifestyle. Our results show that PhoP-PhoQ also plays an essential role in the maintenance of the mutualistic association between *S. glossinidius* and its insect host. In the absence of *phoP*, *S. glossinidius* is highly sensitive to the bactericidal effects of AMPs *in vitro* and is incapable of superinfecting either its natural host or a closely related hippoboscid louse fly.

In *S. enterica* and other facultative pathogens, the sensor kinase PhoQ plays an important role in the sensing of magnesium, AMPs and acidic pH. Thus, the most striking result to emerge from the current study is that the PhoP-PhoQ system of *S. glossinidius* differs from its counterparts in other bacteria by having a substantially reduced ability to sense these environmental cues. For example, the *S. glossinidius* PhoP-PhoQ system elicited only a minor reduction (c. 2-fold) in the expression of target genes in response to a 1000-fold increase in the level of magnesium in the culture medium. Similarly, the *S. glossinidius* PhoP-PhoQ system was found to be largely unresponsive to a change in pH; *S. glossinidius* cells cultured under acidic conditions showed only a 2.6-fold increase in the expression of a PhoP-regulated target gene relative to cells

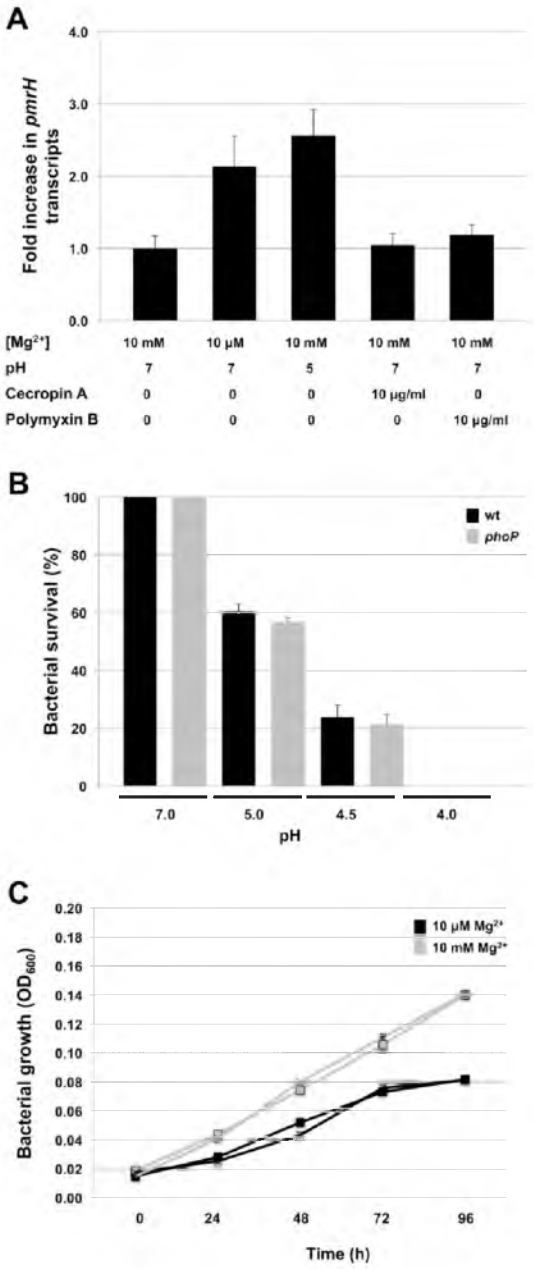


Figure 5. Response of *S. glossinidius* to antimicrobial peptides, acidic pH, or magnesium. A. Quantitative PCR analysis of *pmrH* transcripts derived from *S. glossinidius* cells grown under various medium conditions. Fold increase in *pmrH* transcripts was calculated relative to cells grown in defined medium containing 10 mM of magnesium at pH 7. B. Percent survival of *S. glossinidius* wild type and *phoP* mutant cells at various pH values. C. Growth dynamics of *S. glossinidius* wild type (■) and *phoP* mutant (×) in defined medium containing high (10 mM) and low (10 µM) concentrations of magnesium. Error bars represent standard deviations.
doi:10.1371/journal.pgen.1002349.g005

cultured at neutral pH. In *S. enterica*, the constitutive expression of PhoP-activated genes is known to have a deleterious impact on bacterial virulence and survival in host tissues [43], due to the fact that the initial activation of PhoP induces a transcriptional surge that enables *S. enterica* to rapidly initiate virulence gene expression [44]. In the current study, we were unable to identify any conditions under which the *S. glossinidius* PhoP-PhoQ system was effectively repressed. *In vitro*, *S. glossinidius* was found to display high levels of resistance to AMPs under all conditions tested, indicating that the dynamics of PhoP-based gene regulation in *S. glossinidius* are locked in a constitutively active state, at least with respect to signaling mediated by magnesium, AMPs and pH change.

The presence of a canonical PhoP binding site in the promoter sequence of an inactivated *mgtB* allele indicates that in the recent evolutionary past, *S. glossinidius* used a magnesium-sensing PhoP-PhoQ to control the expression of a high affinity magnesium transporter. The proposed reduction in PhoQ's ability to sense magnesium (and other signals) can be rationalized in several ways. First, if the primary mandate of PhoQ's magnesium sensing capability was to control magnesium uptake, then the inactivation of the gene encoding this magnesium transporter could have facilitated relaxed selection on the ability of PhoQ to sense magnesium. Second, since magnesium binding drives the repression of PhoP-regulated target genes, the loss of a requirement to repress these genes, perhaps resulting from a switch to a static host-associated lifestyle, could also have facilitated relaxed selection on PhoQ's sensing capabilities. Third, it is conceivable that the loss of PhoQ's ability to sense magnesium and acidic pH provided an adaptive advantage in *S. glossinidius*' current lifestyle. Like many recently derived insect endosymbionts, *S. glossinidius* inhabits the hemolymph of its insect host [32] where it is anticipated to be exposed simultaneously to high concentrations of both AMPs [20,25,26] and magnesium [12,45], at neutral pH [46]. In this environment, the magnesium- or pH-mediated repression of PhoP-regulated genes would likely be deleterious due to the fact that it would yield an AMP sensitive phenotype that is incompatible with survival in the AMP-rich hemolymph. Conversely, a magnesium-insensitive PhoQ would be expected to facilitate the constitutive activation of PhoP-regulated target genes, ensuring resistance towards AMPs in magnesium-rich hemolymph.

Since the mechanism of magnesium binding by the *S. enterica* PhoQ homologue has been well characterized through structural and functional studies [38,47], we inspected the sequence of the *S. glossinidius* PhoQ protein to determine if there are any obvious modifications that can explain it's reduced sensitivity to magnesium. Notably, the *S. glossinidius* PhoQ sequence maintains several amino acid substitutions that replace key acidic residues in a location corresponding to a magnesium ligand-binding site in the PhoQ protein of *S. enterica* (Figure S1) [38,47]. In *S. enterica*, mutant strains harboring PhoQ sequences lacking just one of these key acidic residues demonstrate >5-fold reduction in their ability to repress the transcription of PhoP-activated genes in response to magnesium (p*phoQ* D150A in Figure S1 and [38]). Thus it is conceivable that the loss of three key acidic residues in the *S. glossinidius* PhoQ substantially reduced the ability of this sensor kinase to bind magnesium, explaining the reduction in the magnesium dependent PhoQ transcriptional repression of PhoP-activated genes observed in *S. glossinidius*. An alignment of PhoQ sequences derived from a selected range of gamma Proteobacteria shows that PhoQ homologues from two insect endosymbionts (*S. glossinidius* and *Arsenophonus nasoniae*) along with the soft-rot plant pathogen *Dickeya dadantii* (formerly *Erwinia chrysanthemi*) each have putative magnesium-binding patches that lack the requisite acidic

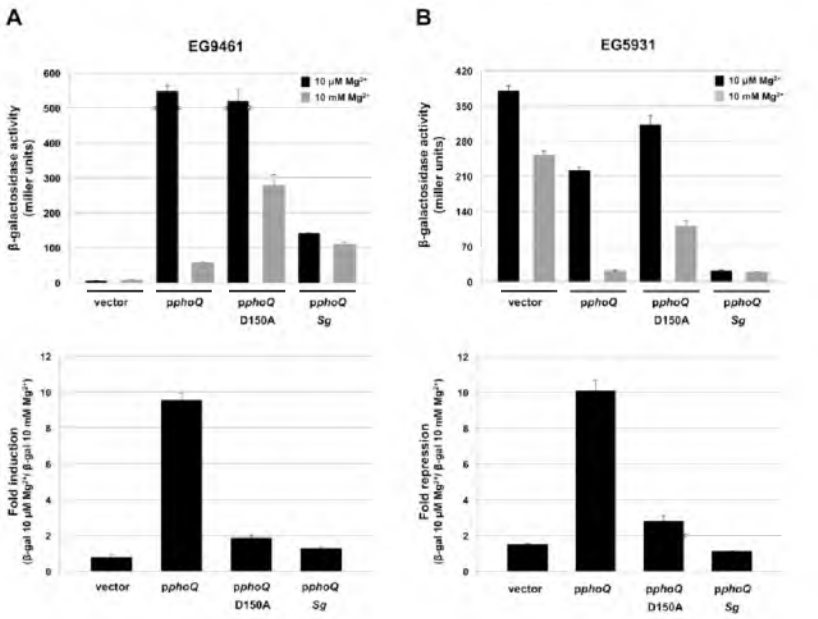


Figure 6. *Salmonella enterica* strains expressing *S. glossinidius* PhoQ do not respond to magnesium. A. β-galactosidase activity (top) of the *S. enterica* PhoP-activated *pmrC* gene produced by strain EG9461 (*phoQ*::Tn*10 pmrC*:: MudJ) harboring the plasmid pUHE21-2 lacI^q (vector) or pUHE21-2 lacI^q derivatives expressing the *S. enterica* PhoQ (p*phoQ*), *S. enterica* PhoQ variant with mutation D150A (p*phoQ* D150A) or *S. glossinidius* PhoQ (p*phoQ Sg*). Fold induction (bottom) of *pmrC*::MudJ β-galactosidase activity produced by strain EG9461 harboring various plasmid constructs in response to magnesium. B. β-galactosidase activity (top) of the *S. enterica* PhoP-activated *pmrC* gene produced by strain EG5931 (*phoP* phoQ*::Tn*10 pmrC*:: MudJ) harboring various plasmid constructs. EG5931 encodes a *phoP* allele (*phoP**) that efficiently phosphorylates from acetyl phosphate (i.e., independently of PhoQ) [39]. Fold repression (bottom) of *pmrC*::MudJ β-galactosidase activity produced by strain EG5931 harboring various plasmid constructs in response to magnesium. Error bars represent standard deviations.
doi:10.1371/journal.pgen.1002349.g006

residues and are therefore predicted to be compromised in terms of magnesium-sensing (Figure S1A). While mutations in this conserved region likely contribute to a decreased magnesium sensing capability, it is possible that changes in other regions of PhoQ also contribute to this phenotype. Indeed, we noticed that the *S. glossinidius* PhoQ homologue also maintains a hydrophobic phenylalanine residue located at position 121, instead of a charged histidine residue that is found in the *S. enterica* PhoQ (Figure S1A). In *S. enterica*, this charged residue has also been shown to be important for magnesium binding and magnesium mediated repression [38].

Although magnesium sensing is recognized as the primary function of PhoQ [14,15], we cannot rule out the possibility that the *S. glossinidius* PhoQ has evolved the ability to sense a novel signal in the insect host. For example, it has recently been shown that the PhoQ protein of *Edwardsiella tarda* has evolved to detect changes in temperature, in addition to magnesium, to coordinate the expression of protein secretion systems that play an important role in virulence [48]. However, it is clear that the *S. glossinidius* PhoQ has lost the ability to sense key signals that play an important role in the functionality of PhoP-PhoQ in many pathogens. The adaptive benefit of PhoP-PhoQ in *S. glossinidius* now appears to be solely due to the ability of this regulatory system to serve as a constitutive driver for the expression of genes that have critical functions in the symbiosis. To this end, the modification or loss of sensory functions in a two-component system is not necessarily paradoxical, as long as an adaptive benefit is realized as a result of the output of the response regulator. Of course a similar functional outcome could be achieved by modulating the promoter sequences of all PhoP-regulated genes to achieve constitutive expression in the absence of PhoP. Although more evolutionary steps are needed to achieve this outcome, it might ultimately be favored by natural selection because it represents a more frugal solution. Indeed, such a transition might already be underway in *S. glossinidius*, evidenced by the fact that the lipid A modification genes, *pagP* and *pmrE*, are expressed independently of PhoP, whereas PhoP is required for the expression of these genes in *S. enterica* [27].

In a wider evolutionary context, it is interesting to note that intact homologues of *phoP-phoQ* and the lipid modification genes are only present in the genome sequences of recently established insect endosymbionts (Table 2). This suggests that the functions of PhoP-PhoQ and the lipid A modification genes are required only as a stopgap in the early stages of symbiotic interactions, to enable bacteria to resist attack from the host immune system to facilitate the establishment and maintenance of persistent infections in host tissues. This is supported by the results of our superinfection experiments, which show that a *phoP* mutant strain of *S. glossinidius* is unable to establish infection in an insect host. Taken together, these observations suggest that recently established insect endosymbionts have an intrinsic ability to overcome the insect immune system and establish persistent infections in insects, and this may help to explain the broad distribution of certain recently established insect endosymbionts (e.g. relatives of *S. glossinidius* and *Arsenophonus* spp.) in distantly related host insects [10,21,42,49 51]. It also explains differences observed in patterns of co-

Table 2. Distribution of *phoP-phoQ*, the magnesium transporters *mgtA* and *mgtB*, and lipid A modification genes among the insect pathogen *Photorhabdus luminescens* and recently derived and ancient insect endosymbionts.

Organism	Classification	Genome size (Mbp)	Age of Association	*phoP*	*phoQ*	*mgtA*	*mgtB*	Lipid A modification genes							
								pagP	*pmrE*	*pmrF*	*pmrH*	*pmrI*	*pmrJ*	*pmrK*	*pmrL*	*pmrM*
Photorhabdus luminescens	Insect pathogen	5.68	N. A.	+	**4**	–	**P**	+	+	+	+	+	+	+	+	+
Arsenophonus nasoniae	Insect symbiont	>3.56	Unknown	+	**2**	–	–	+	+	+	+	+	+	+	+	+
Sodalis glossinidius	Insect symbiont	4.17	Unknown	+	**2**	–	**P**	+	+	+	+	+	+	+	+	+
Candidatus Hamiltonella defensa	Insect symbiont	2.10	Unknown	–	–	+	–	–	+	+	–	+	+	+	+	+
Candidatus Serratia symbiotica	Insect symbiont	2.57	Unknown	+	**1**	–	–	–	+	+	–	–	–	+	+	–
Wigglesworthia glossinidia	Insect symbiont	0.69	40–80	–	–	–	–	–	+	–	–	+	+	+	–	–
Blochmannia spp.	Insect symbiont	0.70–0.79	50	–	–	–	–	-	–	–	–	–	–	–	–	–
Carsonella ruddii	Insect symbiont	0.16	120	–	–	–	–	–	–	–	–	–	–	–	–	–
Buchnera spp.	Insect symbiont	0.42–0.65	150	–	–	–	–	–	–	–	–	–	–	–	–	–
Baumannia cicadellinicola	Insect symbiont	0.68	170	–	–	–	–	–	–	–	–	–	–	–	–	–

Estimated age of association (in million of years) is shown for ancient insect symbionts [58]. "P" indicates that a given sequence was determined to be a pseudogene. "N. A." signifies not applicable. Numbers of acidic residues within the acidic patch are displayed for PhoQ homologues. Note also that *Candidatus* Serratia symbiotica maintains a truncated PhoQ homologue.

doi:10.1371/journal.pgen.1002349.t002

evolution between insects and endosymbionts of different inferred ages. While the phylogenies of insects and their ancient endosymbionts demonstrate high levels of congruence, implying little or no ongoing horizontal symbiont transmission or novel colonization events, the phylogenies of insects and their recently established endosymbionts are often discordant [1].

At a broad level, the results obtained in the current study reinforce the notion that the molecular mechanisms facilitating host-symbiont interactions have a common origin in both pathogenic and mutualistic associations. More specifically, our findings illustrate the capability of a complex regulatory circuit to adapt to a change in lifestyle. Furthermore, the degeneration of the sensing capability of PhoQ may represent a snapshot of a wider picture of regulatory simplification that is concomitant with a transition to a static, mutualistic, host-associated lifestyle that reduces the requirement for bacteria to engage in environmental sensing.

Materials and Methods

Bacterial Strains and Culture Conditions

Escherichia coli W25113 was maintained in Luria-Bertani (LB) medium at 37°C, or 30°C when harboring the repA101 (ts) plasmid pKD46 [52]. *Salmonella enterica* strains were maintained at 37°C in either LB medium or N-minimal medium [38] containing either 10 mM or 10 μM $MgCl_2$. *Sodalis glossinidius* strains were maintained at 25°C in the semi-defined liquid Mitsuhashi-Maramorosch (MM) medium as described previously [42] or in a defined medium composed of 6 g/l casamino acids, 4 g/l glucose, 0.2 g/l KCl, 7.0 g/l NaCl, 0.12 g/l $NaHCO_3$, 0.18 g/l NaH_2PO_3, 10 μg/ml of thiamine, 10 mM or 10 μM of $CaCl_2$, and 10 mM or 10 μM $MgCl_2$, pH 7. When rapidly dividing symbiont cultures were needed, liquid cultures of *S. glossinidius* were maintained in an orbital shaker at 200 rpm. For isolation of symbiont clones, bacteria were plated on 1% MM agar plates, supplemented with 5% defibrinated horse blood, and incubated at 25°C under microaerophilic conditions (5% Oxygen, 10% CO_2, balanced with N_2). Where appropriate, antibiotics were added to culture media at the following concentrations: 100 μg/ml (*E. coli* and *S. enterica*) or 50 μg/ml (*S. glossinidius*) of ampicillin, 15 μg/ml (chromosomal insertions) or 50 μg/ml (high copy number plasmids) of chloramphenicol, 50 μg/ml of kanamycin, 10 μg/ml tetracycline. When needed, 3′-5′-cyclic adenosine monophosphate (cAMP; Sigma Aldrich) was added to the culture medium at a final concentration of 5 mM.

Construction of Replacement Alleles

Replacement alleles for *S. glossinidius* were generated using the lambda-Red recombineering system in *E. coli* [52]. The strategy for construction of replacement alleles for *S. glossinidius* is illustrated in Figure S2. Briefly, *S. glossinidius* genes were amplified by polymerase chain reaction (PCR) using Phusion High-Fidelity DNA polymerase (New England Biolabs). PCR products were ligated into pCR-Blunt II-TOPO (Invitrogen) and transformed into *E. coli* by electroporation. Clones were grown and plasmid DNA was isolated using the QIAprep Spin Miniprep Kit (Qiagen). Recombinant plasmids were then transformed into *E. coli* (pKD46) and replacement alleles were generated by lambda-Red mediated insertion of a chloramphenicol resistance cassette (derived from pEpiFOS-5, Epicentre), according to the method described previously [52]. Following curing of pKD46, plasmids containing the replacement alleles were extracted, diluted and introduced into *E. coli* by electroporation. The replacement alleles were verified by DNA sequencing and plasmids containing the

correct allele sequence were used as PCR templates for the generation of linear DNA substrates for recombineering in *S. glossinidius*.

Transformation of *Sodalis glossinidius*

All *S. glossinidius* transformations (using both plasmid and linear DNA constructs) were conducted using a heat-shock method [53]. Following transformation, the cells were allowed to recover overnight at 25 °C in liquid MM medium prior to plating.

Lambda-Red Mediated Insertions in *Sodalis glossinidius*

Lambda-Red mediated chromosomals insertions were generated as described previously [24]. Briefly, cultures of *S. glossinidius* harboring pKD46 were grown without shaking to an OD_{600nm} of approximately 0.2. The cultures were then transferred to a shaking incubator and grown overnight until OD_{600nm} reached 0.5. The cells were pelleted at 5,000 × *g* for 10 min at 4°C, and washed twice with an equal volume of 0.85% (w/v) NaCl and resuspended in a final volume of 1 ml of 0.85% (w/v) NaCl. The cell suspensions were inoculated into 200 ml of fresh MM liquid medium [42] supplemented with 0.5% (w/v) arabinose and 5 mM cAMP, and incubated at 25°C with shaking for 30 min to facilitate induction of the lambda-Red functions. After induction, the cells were made chemically competent and transformed with linear DNA using the heat-shock method [53]. Following overnight recovery, the cells were resuspended in 150 µl of MM liquid medium. To assess cell viability following transformation, 15 µl of cell suspension was plated on MM blood agar plates alone. To select for recombinant clones, the remaining 135 µl was spread on an MM blood agar plate containing ampicillin and chloramphenicol. Plates were incubated under microaerophilic conditions and inspected for growth after 9 days. Putative recombinant clones were then isolated as single colonies and the presence of chromosomal insertions was confirmed by DNA sequencing.

Curing of Lambda-Red Plasmid from *Sodalis glossinidius*

Following lambda-Red recombineering, plasmid pKD46 was cured from *S. glossinidius* by maintaining recombinant strains in the absence of plasmid selection. Cultures were grown under these conditions with shaking for approximately 50 generations prior to passage into fresh medium. Cultures were passaged a total of five times and then plated on MM blood agar plates supplemented with chloramphenicol alone. After 7 days of growth, colonies were screened for ampicillin sensitivity by replica plating onto MM blood agar plates supplemented with chloramphenicol and ampicillin.

Antimicrobial Peptide Resistance Assay

Antimicrobial peptide resistance assays were performed using a modified version of a previously described method [54]. Wild type and *phoP* mutant strains of *S. glossinidius* were grown to mid-log phase in MM liquid medium. 20 ml of each culture was transferred to 50 ml tubes and the cells were harvested by centrifugation at 5,000 × *g* for 10 minutes at 4°C. The cells were washed twice with an equal volume of 0.85% (w/v) NaCl and resuspended in 1 ml of 0.85% (w/v) NaCl. 0.5 ml of each cell suspension was inoculated into 20 ml of defined medium containing either 10 mM or 10 µM $MgCl_2$ and $CaCl_2$. Following an 8 h incubation (with shaking), the cells were harvested and washed, and 1 ml of cell suspension was inoculated into a 50 ml tube containing 9 ml of either 0.85% (w/v) NaCl alone, or 0.85% (w/v) NaCl supplemented with polymyxin B or cecropin A at a final concentration of 50 µg/ml. The cells were incubated at 25°C for 10 minutes, diluted in 0.85% (w/v) NaCl and approximately 1,000 colony forming units (CFU) were plated on MM blood agar plates. Plates were incubated for 5 days at 25°C under microaerophilic conditions and the resulting colonies were counted. All assays were carried out in triplicate.

Acid Resistance Assay

Wild type and *phoP* mutant strains of *S. glossinidius* were grown to mid-log phase in MM liquid medium. The cells were harvested by centrifugation at 5,000 × *g* for 10 minutes at 4°C and washed twice with an equal volume of 0.85% (w/v) NaCl. Approximately 2×10^8 CFU were inoculated in 10 ml of MM medium at various pH levels (7.0, 5.0, 4.5 and 4.0). Following 1 h incubation, the cells were diluted in 0.85% (w/v) NaCl solution and approximately 1,000 CFU were plated on MM blood agar plates. Plates were incubated for 5 days at 25°C under microaerophilic conditions and the resulting colonies were counted. All assays were carried out in triplicate.

Growth Curves

Wild type and *phoP* mutant strains of *S. glossinidius* were grown to mid-log phase in MM liquid medium. 5 ml of each culture was harvested by centrifugation at 5,000 × g for 10 minutes at 4°C. The cells were then washed twice with an equal volume of defined medium containing 10 µM $MgCl_2$ and inoculated into 10 ml aliquots of defined medium containing either 10 mM or 10 µM $MgCl_2$ and $CaCl_2$. The cultures were maintained at 25°C, and measurements of turbidity (OD_{600nm}) were obtained at 24 h intervals.

Salmonella enterica Complementation Experiments

Salmonella enterica serovar Typhimurium EG9461 [27] and EG5931 [39] harboring pUHE21-2 $lacI^q$ [55] or pUHE21-2 $lacI^q$ derivatives were grown overnight in N-minimal medium containing 10 mM $MgCl_2$. Overnight cultures were diluted 1:100 in fresh N-minimal medium containing 10 mM $MgCl_2$ and allowed to grow for 4 h. The cells were harvested by centrifugation at 8,000 × *g* for 3 min at 4°C, washed twice with 0.85% (w/v) NaCl, and resuspended in N-minimal medium containing either 10 mM or 10 µM $MgCl_2$. Following 3 h of growth, β-galactosidase activity was determined as described by Miller [56].

Total Lipid Extraction and Thin Layer Chromatography

Wild type and *phoP* mutant strains of *S. glossinidius* were grown to mid-log phase in MM liquid medium. 200 ml of each culture was harvested by centrifugation at 5,000 × *g* for 10 minutes at 4°C. The cells were then washed twice with an equal volume of 0.85% (w/v) NaCl. After the second wash, the cells were harvested as described above and resuspended in 5 ml of 0.85% (w/v) NaCl. 2.5 ml of each cell suspension was inoculated into 200 ml of defined media containing either 10 mM or 10 µM $MgCl_2$ and $CaCl_2$, and the cultures were incubated for 8 h with shaking. Following incubation, the cells were harvested by centrifugation at 5,000 × *g* for 10 minutes at 4°C and washed with an equal volume of 0.1% (w/v) ammonium acetate. The cell suspensions were centrifuged as described above and resuspended at a final concentration of 3×10^{11} CFU/ml. 200 µl of each cell suspension was used for total lipid extraction in accordance to the Folch method [57]. Lipid extracts were spotted on a 20×20 cm C_{18} thin layer chromatography plate (Whatman) and developed twice with chloroform: methanol: water (60:30:5, by volume). After chroma-

tography, the plate was allowed to dry and lipids were visualized by iodine staining.

RNA Isolation and Quantitative PCR Analyses

RNA was prepared using the SV Total RNA Isolation System (Promega). Following RNA purification, sample aliquots were treated with Turbo DNase (Ambion) to remove contaminating DNA. RNA samples were then reverse transcribed using the Maxima First Strand cDNA Sysnthesis kit (Fermentas). Quantitation of cDNA was performed in triplicate using the Maxima SYBR Green/ROX qPCR Master Mix (Fermentas). Samples were analyzed using an iCycler iQ Multicolor Real-Time PCR Detection System (Bio-Rad). Relative transcript levels were estimated using the standard curve method, with expression levels normalized against a gene encoding a ribosomal protein (*RplB*) that is expressed constitutively in *S. glossinidius* [8]. A list of the primer sets and respective target genes is presented in Table S1.

Microinjection of *Sodalis glossinidius* into Louse and Tsetse Flies

Glossina morsitans morsitans Westwood tsetse flies were maintained at the Institute of Tropical Medicine (Antwerp, Belgium) as described previously [25]. *Pseudolynchia canariensis* hippoboscid louse flies were maintained on a pigeon colony at the University of Utah. Mid-log phase *S. glossinidius* cells were collected in a 1.5 ml tube by centrifugation at 8,000 *g* for 2 min and resuspended in PBS. Insects were injected into the ventral thorax with $\approx 2\times 10^4$ CFU of the *S. glossinidius phoP* mutant strain, *fliM* mutant strain [24], serving as a positive control for tsetse flies, or *fliM* and wild type strain (serving as a positive controls for louse flies). Microinjections were also performed with heat-killed wild type and *fliM* mutant strains (80°C for 15 min) to serve as negative controls. Insects were sacrificed for DNA isolation at various times post-injection and the resulting DNA samples were screened for the presence of bacteria using PCR primers specific for *S. glossinidius* wild type and the chloramphenicol resistance markers of the mutant strains.

Supporting Information

Figure S1 A. Sequence alignment of PhoQ homologues derived from *S. glossinidius* and related Gammaproteobacteria. The box highlights the PhoQ acidic cluster (acidic residues are shaded in red) that is known to be involved in magnesium binding [17,38,47]. The PhoQ homologues of the insect endosymbionts *S. glossinidius* and *A. nasoniae* and the PhoQ homologue of the soft-rot plant pathogen *D. dadantii* have accumulated non-acidic amino acid substitutions within this cluster, suggesting that these proteins have a reduced ability to bind to magnesium and mediate the repression of PhoP-activated genes. Notably, the S. *glossinidius* PhoQ homologue also has a hydrophobic phenylalanine (shaded in blue) at a charged position (histidine 120) that is required for magnesium binding and magnesium-mediated repression in the PhoQ of *S. enterica* [38]. The alignment was generated using the online MAFFT tool [59]. B. Ribbon representation of the monomeric crystal structure of *S. typhimurium* PhoQ. Red-colored side chains represent acidic residues and gray spheres represent magnesium ions. Missing acidic residues in the *S. glossinidius* PhoQ sequence (D149 to D151) are highlighted. The structural diagram was generated using Pymol software (http://www.pymol.org/).
(TIF)

Figure S2 Schematic illustrating construction of replacement alleles for *S. glossinidius* recombineering. The drug resistance marker (chloramphenicol acetyltransferase, *cat*) was amplified using PCR primers with 5′-flanking sequences that match 35 base target sequences in the *S. glossinidius* chromosome. This PCR product was then integrated into a plasmid borne copy of the target gene, via lambda-Red mediated homologous recombination [52]. After selection for integration, the plasmid was used as PCR template for the synthesis of the replacement construct.
(TIF)

Table S1 Sequences of oligonucleotide primers used in quantitative PCR experiments.
(DOC)

Acknowledgments

We thank Serap Aksoy (Yale University), Eduardo Groisman (Yale University), and Kelly Hughes (University of Utah) for the provision of bacterial strains. We also thank David Blair (University of Utah) for assistance with structural analysis of PhoQ and several anonymous reviewers for useful suggestions.

Author Contributions

Conceived and designed the experiments: MHP KLS JVDA CD. Performed the experiments: MHP KLS LVD JVDA CD. Analyzed the data: MHP KLS CD. Contributed reagents/materials/analysis tools: JVDA CD. Wrote the paper: MHP CD.

References

1. Dale C, Moran NA (2006) Molecular interactions between bacterial symbionts and their hosts. Cell 126: 453 465.
2. Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, et al. (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277: 1453 1474.
3. Chen XA, Li S, Aksoy S (1999) Concordant evolution of a symbiont with its host insect species: Molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. J Mol Evol 48: 49 58.
4. Thomson NR, Howard S, Wren BW, Holden MT, Crossman L, et al. (2006) The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081. PLoS Genet 2: e206. doi:10.1371/journal.pgen.0020206.
5. Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, et al. (2006) Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Res 16: 149 156.
6. Dale C, Young SA, Haydon DT, Welburn SC (2001) The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. Proc Natl Acad Sci U S A 98: 1883 1888.
7. Dale C, Plague GR, Wang B, Ochman H, Moran NA (2002) Type III secretion systems and the evolution of mutualistic endosymbiosis. Proc Natl Acad Sci U S A 99: 12397 12402.
8. Dale C, Jones T, Pontes M (2005) Degenerative evolution and functional diversification of type-III secretion systems in the insect endosymbiont *Sodalis glossinidius*. Mol Biol Evol 22: 758 766.
9. Moran NA, Degnan PH, Santos SR, Dunbar HE, Ochman H (2005) The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. Proc Natl Acad Sci U S A 102: 16919 16926.
10. Nováková E, Hypsa V (2007) A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. FEMS Microbiol Lett 269: 131 135.
11. Degnan PH, Yu Y, Sisneros N, Wing RA, Moran NA (2009) *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. Proc Natl Acad Sci U S A 106: 9063 9068.
12. Derzelle S, Turlin E, Duchaud E, Pages S, Kunst F, et al. (2004) The PhoP-PhoQ two-component regulatory system of *Photorhabdus luminescens* is essential for virulence in insects. J Bacteriol 186: 1270 1279.
13. Groisman EA (2001) The pleiotropic two-component regulatory system PhoP-PhoQ. J Bacteriol 183: 1835 1842.
14. Perez JC, Shin D, Zwir I, Latifi T, Hadley TJ, et al. (2009) Evolution of a bacterial regulon controlling virulence and Mg(2+) homeostasis. PLoS Genet 5: e1000428. doi:10.1371/journal.pgen.1000428.
15. Groisman EA, Mouslim C (2006) Sensing by bacterial regulatory systems in host and non-host environments. Nat Rev Microbiol 4: 705 709.
16. Mitrophanov AY, Jewett MW, Hadley TJ, Groisman EA (2008) Evolution and dynamics of regulatory architectures controlling polymyxin B resistance in enteric bacteria. PLoS Genet 4: e1000233. doi:10.1371/journal.pgen.1000233.

17. Bader MW, Sanowar S, Daley ME, Schneider AR, Cho U, et al. (2005) Recognition of antimicrobial peptides by a bacterial sensor kinase. Cell 122: 461 472.
18. Prost LR, Daley ME, Le Sage V, Bader MW, Le Moual H, et al. (2007) Activation of the bacterial sensor kinase PhoQ by acidic pH. Mol Cell 26: 165 174.
19. Guani-Guerra E, Santos-Mendoza T, Lugo-Reyes SO, Terán LM (2010) Antimicrobial peptides: general overview and clinical implications in human health and disease. Clin Immunol 135: 1 11.
20. Lemaitre B, Hoffmann J (2007) The host defense of *Drosophila melanogaster*. Annu Rev Immunol 25: 697 743.
21. Dale C, Beeton M, Harbison C, Jones T, Pontes M (2006) Isolation, pure culture, and characterization of "*Candidatus* Arsenophonus arthropodicus," an intracellular secondary endosymbiont from the hippoboscid louse fly *Pseudolynchia canariensis*. Appl Environ Microbiol 72: 2997 3004.
22. Haines LR, Hancock RE, Pearson TW (2003) Cationic antimicrobial peptide killing of African trypanosomes and *Sodalis glossinidius*, a bacterial symbiont of the insect vector of sleeping sickness. Vector Borne Zoonotic Dis 3: 175 186.
23. Hao Z, Kasumba I, Lehane MJ, Gibson WC, Kwon J, et al. (2001) Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis. Proc Natl Acad Sci U S A 98: 12648 12653.
24. Pontes MH, Dale C (2011) Lambda Red-mediated genetic modification of the insect endosymbiont *Sodalis glossinidius*. Appl Environ Microbiol 77: 1918 1920.
25. Akoda K, Van den Bossche P, Marcotty T, Kubi C, Coosemans M, et al. (2009) Nutritional stress affects the tsetse fly's immune gene expression. Med Vet Entomol 23: 195 201.
26. Boulanger N, Brun R, Ehret-Sabatier L, Kunz C, Bulet P (2002) Immunepeptides in the defense reactions of *Glossina morsitans* to bacterial and *Trypanosoma brucei brucei* infections. Biochem Mol Biol 32: 369 375.
27. Shi Y, Cromie MJ, Hsu FF, Turk J, Groisman EA (2004) PhoP-regulated Salmonella resistance to the antimicrobial peptides magainin 2 and polymyxin B. Mol Microbiol 53: 229 241.
28. Gunn JS (2008) The *Salmonella* PmrAB regulon: lipopolysaccharide modifications, antimicrobial peptide resistance and more. Trends Microbiol 16: 284 290.
29. Minagawa S, Ogasawara H, Kato A, Yamamoto K, Eguchi Y, et al. (2003) Identification and molecular characterization of the Mg^{2+} stimulon of *Escherichia coli*. J Bacteriol 185: 3696 3702.
30. Zwir I, Shin D, Kato A, Nishino K, Latifi T, et al. (2005) Dissecting the PhoP regulatory network of *Escherichia coli* and *Salmonella enterica*. Proc Natl Acad Sci USA 102: 2862 2867.
31. Bajaj V, Lucas RL, Hwang C, Lee CA (1996) Co-ordinate regulation of *Salmonella typhimurium* invasion genes by environmental and regulatory factors is mediated by control of *hilA* expression. Mol Microbiol 22: 703 714.
32. Cheng Q, Aksoy S (1999) Tissue tropism, transmission and expression of foreign genes in vivo in midgut symbionts of tsetse flies. Insect Mol Biol 8: 125 132.
33. Pontes MH, Dale C (2006) Culture and manipulation of insect facultative symbionts. Trends Microbiol 14: 406 412.
34. Weiss BL, Wu Y, Schwank JJ, Tolwinski NS, Aksoy S (2008) An insect symbiosis is influenced by bacterium-specific polymorphisms in outer-membrane protein A. Proc Natl Acad Sci USA 105: 15088 15093.
35. Weiss BL, Wang J, Aksoy S (2011) Tsetse immune system maturation requires the presence of obligate symbionts in larvae. PLoS Biol 9: e1000619. doi:10.1371/journal.pbio.1000619.
36. Petersen FT, Meier R, Kutty SN, Wiegmann BM (2007) The phylogeny and evolution of host choice in the Hippoboscoidea (Diptera) as reconstructed using four molecular markers. Mol Phylogenet Evol 45: 111 122.
37. Garcia Véscovi E, Soncini FC, Groisman EA (1996) Mg^{2+} as an extracellular signal: environmental regulation of Salmonella virulence. Cell 84: 165 174.
38. Chamnongpol S, Cromie M, Groisman EA (2003) Mg^{2+} sensing by the Mg^{2+} sensor PhoQ of *Salmonella enterica*. J Mol Biol 325: 795 807.
39. Chamnongpol S, Groisman EA (2000) Acetyl phosphate-dependent activation of a mutant PhoP response regulator that functions independently of its cognate sensor kinase. J Mol Biol 300: 291 305.
40. Snavely MD, Miller CG, Maguire ME (1991) The *mgtB* Mg^{2+} transport locus of *Salmonella typhimurium* encodes a P-type ATPase. J Biol Chem 266: 815 823.
41. Tao T, Snavely MD, Farr SG, Maguire ME (1995) Magnesium transport in Salmonella *typhimurium*: *mgtA* encodes a P-type ATPase and is regulated by Mg^{2+} in a manner similar to that of the *mgtB* P-type ATPase. J Bacteriol 177: 2654 2662.
42. Dale C, Maudlin I (1999) *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. Int J Syst Bacteriol 49: 267 275.
43. Miller SI, Mekalanos JJ (1990) Constitutive expression of the phoP regulon attenuates *Salmonella* virulence and survival within macrophages. J Bacteriol. 172: 2485 2490.
44. Shin D, Lee EJ, Huang H, Groisman EA (2006) A positive feedback loop promotes transcription surge that jump-starts Salmonella virulence circuit. Science 314: 1607 1609.
45. Natochin YV, Parnova RG (1987) Osmolality and electrolyte concentration of hemolymph and the problem of ion and volume regulation of cells in higher insects. Comp Biochem Physiol A 88: 563 570.
46. Petit JP (1968) Hemolymph of *Glossina*: collection and analysis. Rev Elev Med Vet Pays Trop 21: 493 498.
47. Cho US, Bader MW, Amaya MF, Daley ME, Klevit RE, et al. (2006) Metal bridges between the PhoQ sensor domain and the membrane regulate transmembrane signaling. J Mol Biol 356: 1193 1206.
48. Chakraborty S, Li M, Chatterjee C, Sivaraman J, Leung KY, et al. (2010) Temperature and Mg^{2+} sensing by a novel PhoP-PhoQ two-component system for regulation of virulence in *Edwardsiella tarda*. J Biol Chem 285: 38876 38888.
49. Fukatsu T, Koga R, Smith WA, Tanaka K, Nikoh N, et al. (2007) Bacterial endosymbiont of the slender pigeon louse, *Columbicola columbae*, allied to endosymbionts of grain weevils and tsetse flies. Appl Environ Microbiol 73: 6660 6668.
50. Kaiwa N, Hosokawa T, Kikuchi Y, Nikoh N, Meng XY, et al. (2010) Primary gut symbiont and secondary, *Sodalis*-allied symbiont of the Scutellerid stinkbug *Cantao ocellatus*. Appl Environ Microbiol 76: 3486 3494.
51. Nováková E, Hypsa V, Moran NA (2009) *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. BMC Microbiol 9: 143.
52. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97: 6640 6645.
53. Inoue H, Nojima H, Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. Gene 96: 23 28.
54. Groisman EA, Heffron F, Solomon F (1992) Molecular genetic analysis of the *Escherichia coli phoP* locus. J Bacteriol 174: 486 491.
55. Soncini FC, Vescovi EG, Groisman EA (1995) Transcriptional autoregulation of the *Salmonella typhimurium phoPQ* operon. J Bacteriol 177: 4364 4371.
56. Miller JH (1992) A short course in bacterial genetics: a laboratory manual and handbook for Escherichia coli and related bacteria. New York: Cold Spring Harbor Laboratory Press. pp 456.
57. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D (2008) Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. J Lipid Res 49: 1137 1146.
58. Gosalbes MJ, Latorre A, Lamelas A, Moya A (2010) Genomics of intracellular symbionts in insects. Int J Med Microbiol 300: 271 278.
59. Katoh, K, Misawa, K, Kuma, K, Miyata, T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acid Res 30: 3059 3066.

APPENDIX B

QUORUM SENSING PRIMES THE OXIDATIVE STRESS RESPONSE IN THE INSECT ENDOSYMBIONT, SODALIS GLOSSINIDIUS

Reprinted with permission from:
Pontes, M. H., M. Babst, R. Lochhead, K. Oakeson, K. Smith, and C. Dale. 2008. Quorum sensing primes the oxidative stress response in the insect endosymbiont, Sodalis glossinidius. *PLoS ONE*, 3:e3541.

OPEN ACCESS Freely available online

PLoS one

Quorum Sensing Primes the Oxidative Stress Response in the Insect Endosymbiont, *Sodalis glossinidius*

Mauricio H. Pontes*, Markus Babst, Robert Lochhead, Kelly Oakeson, Kari Smith, Colin Dale

Department of Biology, University of Utah, Salt Lake City, Utah, United States of America

Abstract

Background: *Sodalis glossinidius*, a maternally transmitted bacterial endosymbiont of tsetse flies (Glossina spp.), uses an acylated homoserine lactone (AHL)-based quorum sensing system to modulate gene expression in accordance with bacterial cell density. The *S. glossinidius* quorum sensing system relies on the function of two regulatory proteins; SogI (a LuxI homolog) synthesizes a signaling molecule, characterized as N-(3-oxohexanoyl) homoserine lactone (OHHL), and SogR1 (a LuxR homolog) interacts with OHHL to modulate transcription of specific target genes.

Methodology/Principal Findings: We used a tiling microarray to analyze the *S. glossinidius* transcriptome in the presence and absence of exogenous OHHL. The major finding is that OHHL increases transcription of a large number of genes that are known to be involved in the oxidative stress response. We also show that the obligate symbiont of the rice weevil, *Sitophilus oryzae* (SOPE), maintains copies of the quorum sensing regulatory genes that are found in *S. glossinidius*. Molecular evolutionary analyses indicate that these sequences are evolving under stabilizing selection, consistent with the maintenance of their functions in the SOPE symbiosis. Finally, the expression studies in *S. glossinidius* also reveal that quorum sensing regulates the expression of a cryptic, degenerate gene (*carA*) that arose from an ancient deletion in the last common ancestor of *S. glossinidius* and SOPE.

Conclusions/Significance: This oxidative stress response is likely mandated under conditions of dense intracellular symbiont infection, when intense metabolic activity is expected to generate a heavy oxidative burden. Such conditions are known to arise in the bacteriocytes of grain weevils, which harbor dense intracellular infections of symbiotic bacteria that are closely related to *S. glossinidius*. The presence of a degenerate *carA* sequence in *S. glossinidius* and SOPE indicates the potential for neofunctionalization to occur during the process of genome degeneration.

Citation: Pontes MH, Babst M, Lochhead R, Oakeson K, Smith K, et al. (2008) Quorum Sensing Primes the Oxidative Stress Response in the Insect Endosymbiont, *Sodalis glossinidius*. PLoS ONE 3(10): e3541. doi:10.1371/journal.pone.0003541

Editor: Raphael H. Valdivia, Duke University Medical Center, United States of America

Received August 20, 2008; **Accepted** October 6, 2008; **Published** October 28, 2008

Funding: Research in our laboratory is supported by a National Science Foundation award EF-0523818 to CD.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: pontes.biology@utah.edu

Introduction

All living organisms are dependent on their ability to sense the physical properties of their environment and respond accordingly through behavioral changes. In the context of symbiosis, the establishment and maintenance of a symbiotic relationship is dependent on the ability of both parties to perceive one another and coordinate their activities. In several recent studies it has been shown that secreted chemicals provide a means for communication between symbiotic partners. For example, chemical communication between legumes and rhizobia is known to mediate many important symbiotic interactions. These include initiation of nodulation [1], evasion of legume defenses [2 5], elongation of infection threads [6,7] and development of nitrogen fixing bacteroids [4,5,8]. The marine symbiosis between *Vibrio fischeri* and the squid *Euprymna scolopes* is also dependent on chemical communication. In this system, chemical signaling mediates the induction of mucus secretion by the squid [9], bacterial attachment and aggregation to the host mucus [10,11], bacterial migration to the squid light organ [12], bacterial light emission [13], and the induction of physiological [14] and developmental [15] changes in the infected squid. Models systems, such as the plant-rhizobia and squid-*Vibrio* symbioses have provided a high level of insight into the molecular processes underlying interactions between symbiotic partners. However, this knowledge is lacking for the majority of symbiotic systems, where sophisticated experimental approaches are not applicable.

Insects from many different taxonomic groups are known to maintain beneficial associations with maternally transmitted bacterial symbionts [16]. Many of these associations are ancient in origin and obligate in nature; in these cases bacterial symbionts often provide their insect hosts with essential nutrients that are lacking in the host's specialized diet [17]. More recently derived associations are typically facultative in nature; in these cases bacterial symbionts often provide ancillary benefits such as protection from natural enemies or enhanced tolerance towards conditional environmental stresses [18]. Regardless of the nature of the association, success in the symbiosis depends upon a complex interplay between bacterial symbionts and their insect hosts throughout the course of insect development and reproduction. For example, many symbiotic bacteria preferentially colonize specialized insect tissues or cells, where they reach extremely high infection densities. This feature is most pronounced in obligate symbionts, which are often found exclusively in specialized organs

called bacteriomes [19]. However, facultative symbionts also tend to colonize specific host cell and tissue types at high infection densities [20,21,22]. In both cases, it seems likely that specific colonization patterns provide physiological benefits for both partners in the association.

In this study we describe the function of a quorum sensing system in the maternally transmitted symbiont of tsetse flies, *Sodalis glossinidius*. Quorum sensing is a chemical signaling process that allows bacteria to monitor their local population density and induce changes in behavior through the modulation of gene expression [23]. Our experiments demonstrate that *S. glossinidius* utilizes an acylated homoserine lactone (AHL)-based quorum sensing system to prime the expression of genes that ameliorate the deleterious effects of oxidative stress that are encountered at high densities of symbiont infection in host tissues.

Results

Characterization of the Quorum Sensing Machinery of *S. glossinidius*

The whole genome sequence of *S. glossinidius* was found to contain a single candidate AHL synthase gene (GenBank locus tag SG0284; designated *sogI* in this report) and two candidate AHL-binding regulatory genes (SG0285 and SG1740, designated *sogR1* and *sogR2* respectively). Whereas *sogI* and *sogR1* are located adjacent to one another, *sogR2* is found at a distant location in the *S. glossinidius* chromosome. To assess the functionality of the *S. glossinidius* quorum sensing system we performed thin layer chromatography (TLC) overlay bioassays with culture supernatant extracts from *S. glossinidius* and a recombinant *E. coli* strain expressing the *S. glossinidius* SogI protein. The bioassay, performed using both *Agrobacterium tumefaciens* [24] and *Chromobacterium violaceum* (data not shown) [25] reporter strains, indicated the presence of a single AHL molecule in the supernatant extracts from both *S. glossinidius* and the recombinant *E. coli* strain (Figure 1A).

To avoid complications associated with the purification of small molecules from the complex *S. glossinidius* culture medium, we elected to perform mass spectrometric characterization of the *S. glossinidius* AHL on the culture supernatant extracted from the recombinant *E. coli* strain expressing the *S. glossinidius* SogI protein. An ethyl acetate extract of culture supernatant was fractionated on a C_{18} column by reverse-phase high performance liquid chromatography. The fractions containing the putative AHL molecule were then identified using the *A. tumefaciens* reporter strain [24] and analyzed by electrospray ionization and collision induced dissociation mass spectrometry (CID-MS). The fragmentation pattern of the candidate AHL ion obtained following CID-MS was consistent with the structure of N-(3-oxohexanoyl) homoserine lactone (OHHL; Figure 1B) [26]. This result is supported by the fact that the AHL molecules produced by *S. glossinidius* and the recombinant *E. coli* strain co-migrated with synthetic and natural forms of OHHL produced by *Pseudomonas aeruginosa* strain PAO1 in our TLC assays (Figure 1A) [27].

Interactions Between OHHL and the SogR1 and SogR2 Transcriptional Regulators

In many bacterial species that utilize AHL quorum sensing systems, the transcription of the *luxI* homolog is autoregulated by quorum sensing [13,28,29,30]. In addition, many quorum-sensing regulated genes contain an 18 to 20-bp long imperfect palindromic sequence upstream of their initiator codon. This inverted repeat sequence, known as *lux* box, serves as the binding site for LuxR transcriptional response regulators [31]. The presence of a canonical *lux* box (ACCTGAACTTTAGTACAGGT) 86 bp upstream from the *S. glossinidius sogI* start codon suggested that the transcription of this gene is autoregulated by quorum sensing. To test this hypothesis and characterize interactions between OHHL and the SogR1 and SogR2 transcriptional regulators, three recombinant plasmids were constructed (pMP2-4; Supplementary Table S1 and Figure 2A). Each of these plasmids harbors an active *sogI-lacZ* reporter fusion, and in addition pMP3 and pMP4 have *sogR1* and *sogR2* expressed from their native promoters. When these plasmids were transformed into isogenic strains of *E. coli*, only cells harboring pMP3 (expressing SogR1) demonstrated increased β-galactosidase activity when synthetic OHHL was added to the culture medium (Figure 2A). This shows

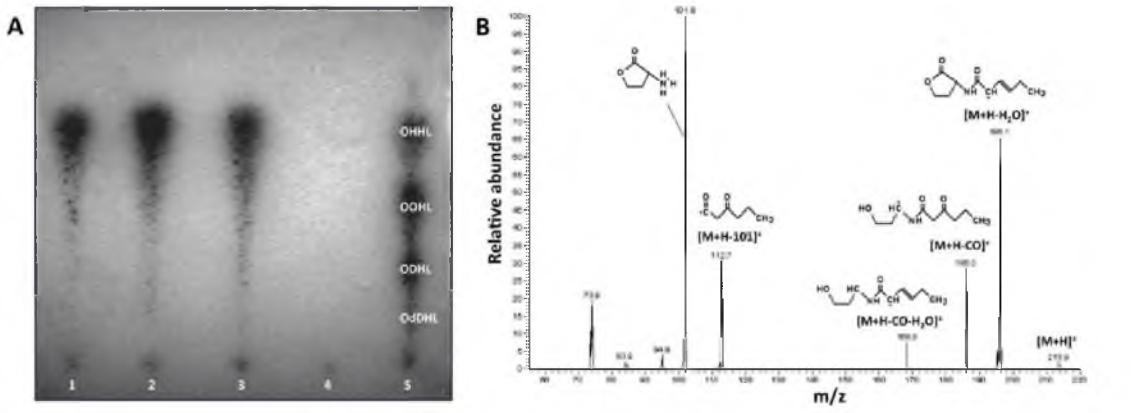


Figure 1. Characterization of *S. glossinidius* AHL. A. AHL samples were separated by TLC and overlaid with a live culture of an *A. tumefaciens* AHL reporter strain. Dark spots indicate the presence of AHL molecules. Lane 1: synthetic N-(3-oxohexanoyl)-homoserine lactone (OHHL); lane 2: *S. glossinidius* culture supernatant extract; lane 3: culture supernatant extract from recombinant *E. coli* expressing *S. glossinidius* SogI; lane 4: culture supernatant extract from an isogenic *E. coli* strain lacking SogI; lane 5: culture supernatant extract from *P. aeruginosa* PAO1 known to produce OHHL, N-(3-oxooctanoyl)-AHL (OOHL), N-3-(oxodecanoyl)-AHL (ODHL) and N-3-(oxododecanoyl)-AHL (OdDHL). B. Mass spectrum of the *S. glossinidius* candidate AHL ion following collision induced dissociation. The fragmentation pattern of this molecule is consistent with the structure of OHHL.
doi:10.1371/journal.pone.0003541.g001

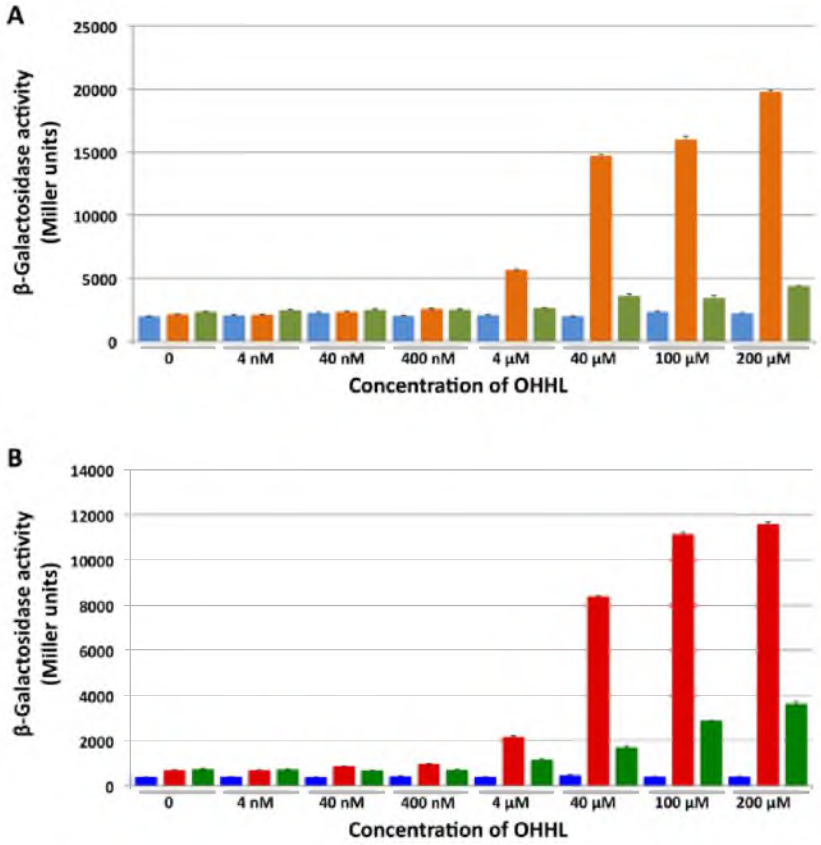


Figure 2. Interactions of *S. glossinidius* SogR-OHHL Complexes with *sogI* and *carA* Promoters. A. β-galactosidase activity of isogenic strains of *E. coli* harboring pMP2 (light blue), pMP3 (orange), pMP4 (light green) following exposure to different concentrations of exogenous OHHL. B. β-galactosidase activity of strains of *E. coli* harboring pMP5 (dark blue), pMP6 (red) or pMP7 (dark green) following exposure to different concentrations of OHHL. Bars indicate standard deviation.
doi:10.1371/journal.pone.0003541.g002

that the interaction between OHHL and the *S. glossinidius* SogR1 protein facilitates the binding of the SogR1-OHHL complex to the *sogI* promoter, resulting in an increase in the rate of transcription of *sogI* (Figure 2A). It should also be noted that cells harboring pMP4 (expressing SogR2) demonstrated no significant increase in β-galactosidase activity when synthetic OHHL was added to the culture medium. This is likely due to a lack of interaction between SogR2 and OHHL and/or the *sogI* promoter.

Identification of Quorum Sensing Regulated Genes in *S. glossinidius*

The *S. glossinidius* tiling microarray was used to identify genes (including putative pseudogenes) that were differentially expressed in the presence and absence of exogenous OHHL. Data was obtained from four replicate microarray experiments and analyzed using a Bayesian analysis of posterior probability. Quantitative PCR (qPCR) assays were then performed to validate the results obtained from the microarray experiments. The data obtained from the qPCR assays closely matched the data obtained from the microarray experiment. The complete dataset obtained from the microarray and qPCR experiments is presented in Supplementary Table S2.

The microarray and qPCR experiments demonstrated that expression of the *S. glossinidius sogI* gene (SG0284) was increased (4.6-fold and 7.5-fold, respectively) in response to the addition of exogenous OHHL. These results are in agreement with those results obtained from the promoter-probe experiment (Figure 2B), validating the experimental approaches used in the microarray and qPCR experiments. Only two other candidate genes (SG0585 and SG0586) displayed a >5-fold increase or decrease in expression in response to OHHL.

According to the microarray data, a substantial number of candidate genes show a statistically significant increase or decrease in expression in the range of 1.2 to 5-fold. In order to understand the biological significance of these changes in gene expression, the data was analyzed according to the clusters of orthologous groups (COG) classification (Figure 3). This revealed a bias in the representation of genes showing >1.2-fold increase in expression in response to the addition of OHHL within COG categories C (energy production and conversion), L (DNA replication, recombination and repair), O (posttranslational modification, protein turnover, chaperones) and P (inorganic ion transport and metabolism). Furthermore, a bias was also detected among those genes showing >1.2-fold decrease in expression in response to the

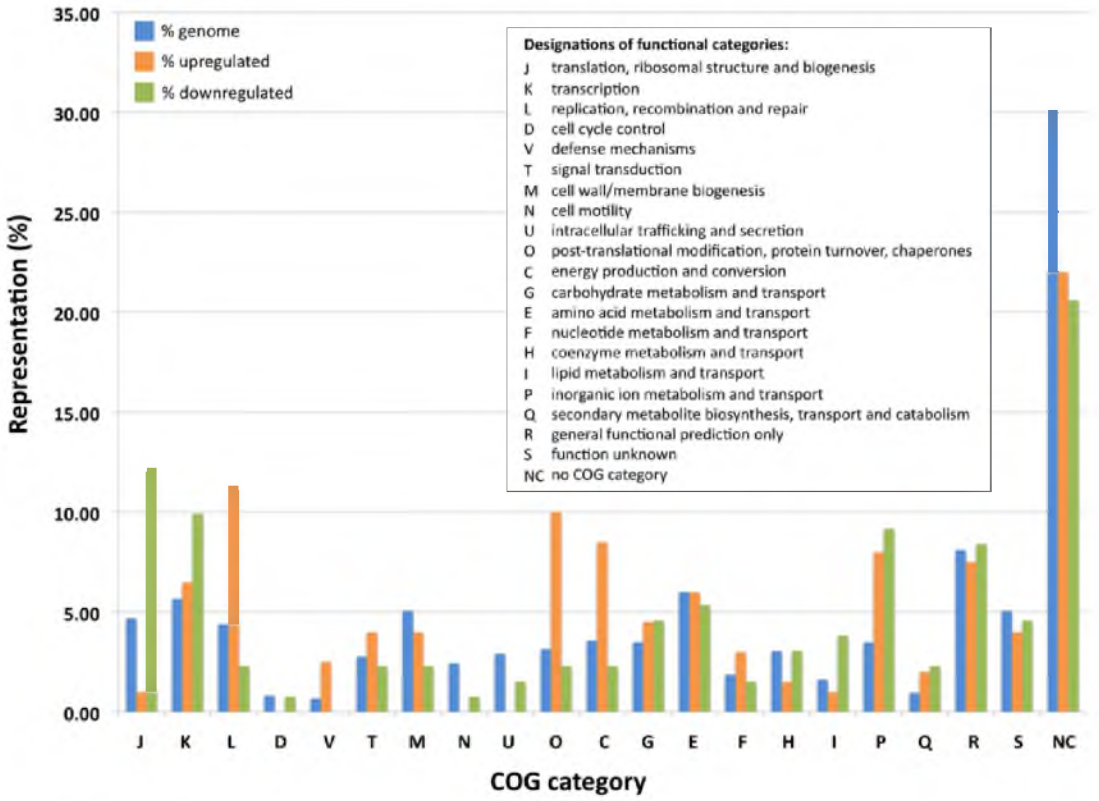


Figure 3. COG-Based Analysis of Microarray Expression Data. Distribution of *S. glossinidius* genes among different COG categories. Blue bars correspond to all protein coding genes in the *S. glossinidius* genome, whereas orange and green bars correspond to genes whose expression increased or decreased (respectively) by at least 1.2-fold in response to addition of OHHL. The COG classifications are described in the inset.
doi:10.1371/journal.pone.0003541.g003

addition of OHHL. The downregulated genes were overrepresented in COG categories J (translation), K (transcription), P and I (lipid transport and metabolism).

Many of the genes that were upregulated in response to OHHL are predicted to play a role in the oxidative stress response. These include genes predicted to encode proteins involved in the breakdown of reactive oxygen species (ROS; e.g., SG0017, SG0642, SG1609 and SG2101), repair of oxidatively damaged cellular components (SG1106 and SG1348), transport of iron and manganese (SG1516, SG1517, SG1518 and SG1519) and protein folding (e.g., SG0409, SG0584, SG0692 and SG2325; Supplementary Table S2). The addition of OHHL also increased the expression of bacterioferritin (SG2280) and genes involved in iron siderophore biosynthesis (SGP1_0041 46) [32]. Based on the fact that the siderophore biosynthetic genes of *Erwinia chrysanthemi* and *S. glossinidius* share high levels of sequence identity, it seems likely that *S. glossinidius* synthesizes an achromobactin-like, citrate-based siderophore [32]. With this in mind, it should be noted that the addition of OHHL stimulated an increase in transcription of a number of citric acid cycle enzymes, including citrate synthase (SG0871), aconitase (SG0477), isocitrate dehydrogenase (SG0700), α-ketoglutarate dehydrogenase (SG0876, SG0877), succinyl-CoA synthase (SG0878, SG0879) and succinate dehydrogenase (SG0872 75; Supplementary Table S2). It is possible that the resulting increase in TCA cycle activity might be required to provide sufficient citric acid for siderophore biosynthesis at high cell density.

Among those genes showing decreased expression in the presence of OHHL, the largest representational bias was found in COG category J (Figure 3). This includes several genes encoding subunits of the 30S (SG0380, SG0412 and SG2269) and 50S ribosomal proteins (SG0133, SG1420, SG1421, SG1572, SG2207, SG2252, SG2270, SG2271 and SG2273). In addition, genes encoding a 16S rRNA pseudouridylate synthase A (SG1570), a tRNA/rRNA methyltransferase (SG1908) and a putative ribosome modulation factor (SG1025) also displayed reduced expression in the presence of OHHL. This leads to an interesting hypothesis perhaps quorum sensing restricts the growth rate of *S. glossinidius* at high infection density by reducing the activities of enzymes involved in translation. However, we were unable to detect any difference in the growth rate of laboratory cultures of *S. glossinidius* in the presence and absence of exogenous OHHL (data not shown).

The *S. glossinidius* gene showing the largest reduction in expression in response to OHHL is annotated as a dethiobiotin synthase (*bioD*; SG1466). *Sodalis glossinidius* is unusual because it maintains two phylogenetically distinct ORFs (SG0906 and SG1466) encoding putative *bioD* homologs. SG0906, whose expression is not affected by OHHL, is located within an operon alongside other genes known to be involved in biotin biosynthesis

(*bioABFC*; SG0902 SG0905). Thus, SG0906 appears to be a component of the canonical biotin biosynthesis gene cluster found in a wide range of bacteria [33]. SG1466 is also unusual because the C-terminal end of the predicted translation product maintains a sugar transporter domain that is not found in any of the other dethiobiotin synthetases in the GenBank database. Furthermore, SG1466 is not located within a cluster of genes involved in biotin biosynthesis. Thus, SG1466 may have evolved to provide a novel function in *S. glossinidius*.

Biochemical Detection of Siderophores

Iron siderophore assays were performed to quantitate siderophore production in *S. glossinidius* cultures maintained at low cell density in the presence and absence of exogenous OHHL. The addition of OHHL provides an "artificial" quorum that stimulates a 5-fold increase in siderophore production in culture media over the course of 24 h (Figure 4). After 48 h, this difference is reduced to 3-fold as the uninduced cultures reach a cell density sufficient to provide a "natural" quorum. These results demonstrate that *S. glossinidius* increases production of iron siderophores in a cell density-dependent manner in response to the quorum sensing signaling molecule, OHHL. This is in agreement with the microarray and qPCR data showing that *S. glossinidius* increases transcription of genes involved in siderophore biosynthesis in response to the addition of OHHL.

Degeneration of Carbapenem Biosynthesis Genes

According to the microarray data, two candidate genes (SG585 and SG586) demonstrated a >5-fold increase in transcription in response to the addition of OHHL. The GenBank annotation of the *S. glossinidius* genome sequence indicates that SG585 and SG586 are genic components of a biosynthetic pathway that leads to the production of a β-lactam antibiotic called carbapenem. Carbapenem is produced by a number of Gram-negative bacteria including *Erwinia* spp., which is one of the closest free-living relatives of *S. glossinidius* [34 37]. Furthermore, carbapenem production is often controlled by quorum sensing, and in *Erwinia carotovora* the signaling molecule is known to be OHHL [38]. The genes involved in the biosynthesis of carbepenem (*carABCDE*) are

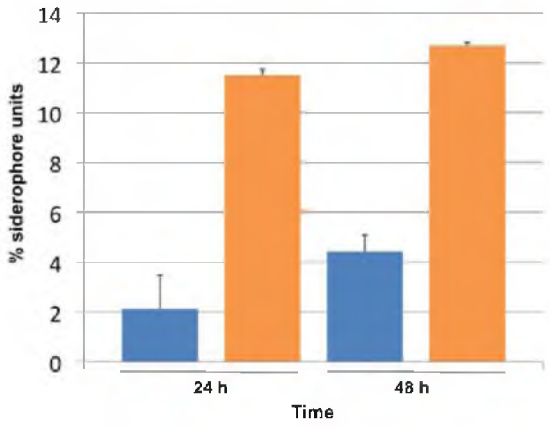


Figure 4. Influence of OHHL on Iron Siderophore Production in *S. glossinidius*. Siderophore units were estimated from samples of *S. glossinidius* culture supernatant in the presence (orange) and absence (blue) of exogenous OHHL.
doi:10.1371/journal.pone.0003541.g004

normally localized in a cluster alongside genes that are known to confer a carbapenem resistance phenotype (*carFG*; Figure 5) [36]. In *S. glossinidius*, most of the genes required for carbapenem biosynthesis and resistance have been lost as a result of a deletion between the C-terminal domain of CarA (SG0586) and the N-terminal domain of CarG (SG0585). Therefore, only the N-terminal domain of CarA and the C-terminal domain of CarG remain. However, several lines of evidence indicate that the genic remnant of *carA* retains functionality. First, it should be noted that the weevil symbiont, SOPE, also maintains truncated copies of *carA* and *carG* that share an almost identical intergenic deletion (Figure 5). Although it is possible that such deletions occurred independently in the lineages leading to *S. glossinidius* and SOPE, it seems more likely (based on parsimony) that a single deletion took place in the common ancestor of these symbionts. Furthermore, if the *carA* genes of *S. glossinidius* and SOPE were inactive, we would anticipate a high ratio of nonsynonymous (nonsilent) to synonymous (silent) substitutions (dN/dS). However, our molecular evolutionary analysis (detailed below) shows that *carA* has a low dN/dS ratio, compatible with the preservation of gene function under stabilizing selection. Second, the truncation in the *S. glossinidius* and SOPE *carA* sequences occurs at a point in the conceptual CarA protein that links an N-terminal nucleophilic hydrolase domain and a C-terminal synthase domain (Figure 5). Thus, the *carA* sequences in *S. glossinidius* and SOPE retain only the nucleophilic hydrolase domain of the ancestral *carA*. Third, the conceptual nucleophilic hydrolase domain in the *S. glossinidius* and SOPE *carA* sequences retain many of the key residues known to be important for amino acid amidohydrolysis (data not shown). Thus, we postulate that the truncated *carA* sequences in *S. glossinidius* and SOPE have acquired a novel functionality, unrelated to antibiotic biosynthesis, as a result of deletion.

Interestingly, *E. carotovora* is known to maintain at least two homologs of LuxR, one of which (CarR) is dedicated to the regulation of the carbapenem gene cluster [39,40]. Since *S. glossinidius* is also known to maintain two LuxR homologs, we decided to investigate interactions between SogR1, SogR2 and the *carA* promoter. Plasmids pMP5 7 were constructed by replacing the *sogI-lacZ* fusions from pMP2 4, respectively, with a *carA-lacZ* fusion. *Escherichia coli* strains harboring these plasmids were then tested for β-galactosidase activity in the presence of increasing concentrations of OHHL (Figure 2B). Interestingly, only cells harboring pMP6 (expressing SogR1) demonstrated increased β-galactosidase activity in response to OHHL. This shows that the canonical SogR1 protein is involved in the cell density dependent regulation of *carA*, and that SogR2 is therefore not a functional homolog of CarR.

Molecular Evolutionary Analyses

In order to understand the molecular evolutionary genetics of the quorum sensing system in *S. glossinidius*, we obtained homologs of *sogI*, *sogR1*, *sogR2* and the truncated *carA* ORF from an unfinished (6 8× coverage) genome sequence of the *Sitophilus oryzae* primary endosymbiont (SOPE), which is known to be closely related to *S. glossinidius* [41,42]. The putative coding sequences of *sogI*, *sogR1* and *sogR2* were found to be intact in SOPE, indicating their potential to encode proteins that serve as regulators of quorum sensing in the weevil symbiosis.

The *luxI* and *luxR* homologs from *S. glossinidius*, SOPE and other Gram negative bacteria were used to construct maximum likelihood (ML) phylogenetic trees. Both the *luxI* and *luxR* trees were supported by more than 50% of ML bootstrap resamples at every node. Four major clades were resolved in the *luxI* tree (Figure 6A), each with >95% bootstrap support. The *luxI*

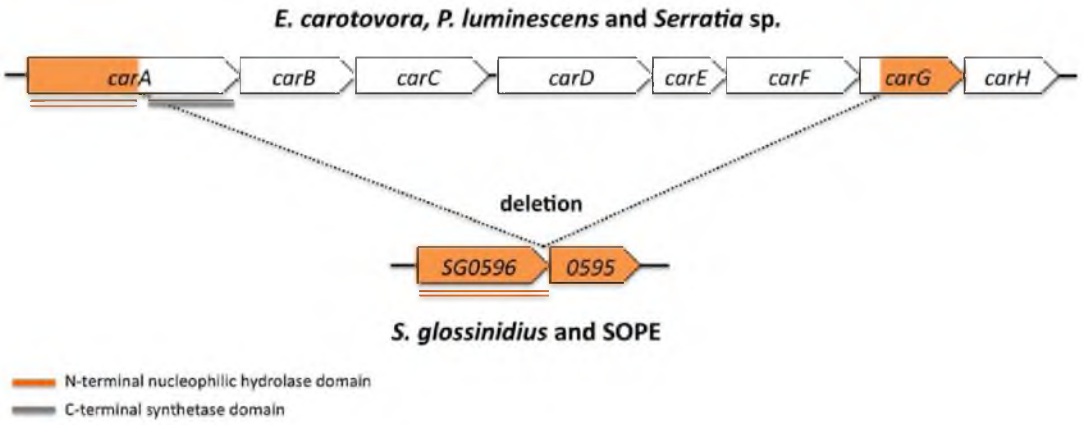


Figure 5. Degeneration of a Carbapenem Biosynthesis Gene Cluster in *S. glossinidius* and SOPE. Carbapenem production (*carABCDE*) and resistance (*carFG*) genes are clustered in the genomes of *E. carotovora*, *P. luminescens* and *Serratia* sp. *Sodalis glossinidius* and SOPE maintain truncated copies of *carA* and *carG* that likely arose through the deletion of intervening DNA.
doi:10.1371/journal.pone.0003541.g005

homologs from *S. glossinidius* and SOPE were placed in a clade with 99% bootstrap support alongside *luxI* homologs from *Erwinia* spp. and *Yersinia* spp., which are the closest free-living relatives of *S. glossinidius* [43]. Furthermore, the *S. glossinidius* and SOPE sequences were placed together in a sub-clade with 100% bootstrap support, indicating their close ancestry. The same pattern of relationships was resolved in the *luxR* tree (Figure 6B). Both the *luxR* homologs from *S. glossinidius* and SOPE were placed in a clade along with *luxR* homologs from *Erwinia* spp. and *Yersinia* spp. with 97% bootstrap support. Interestingly, the *S. glossindius*

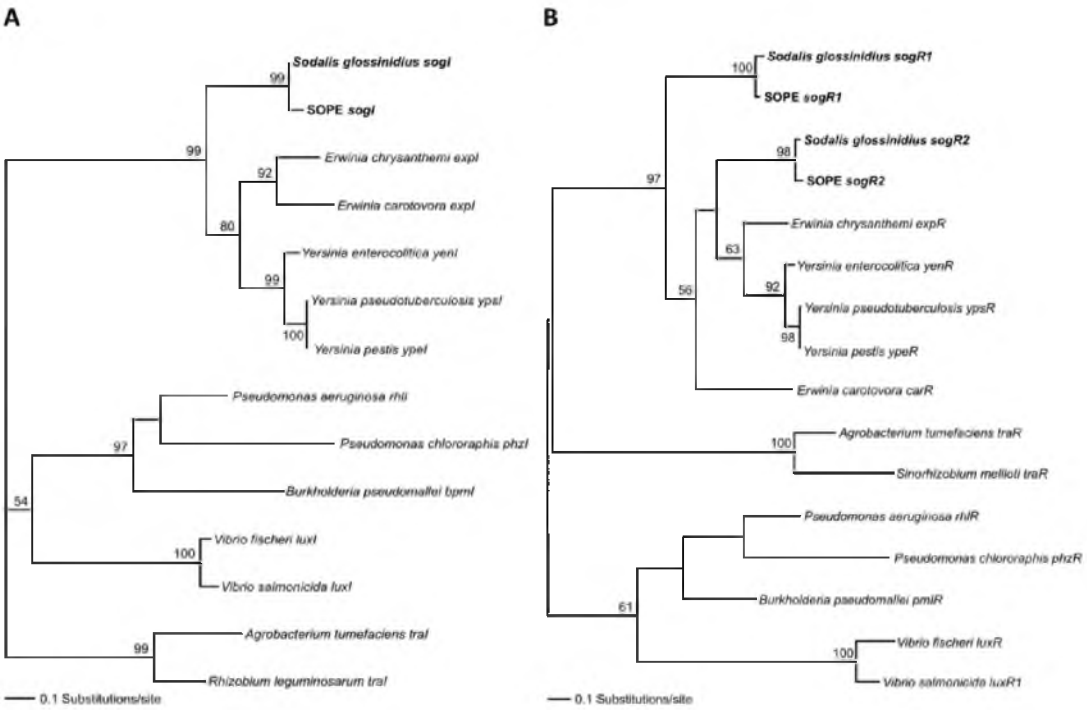


Figure 6. Common Ancestry of *S. glossinidius* and SOPE Quorum Sensing Regulatory Genes. Maximum likelihood phylogenetic analyses provide support for the common ancestry of the *sogI* (A), *sogR1* and *sogR2* (B) sequences in *S. glossinidius* and SOPE. Bootstrap values >50% are shown adjacent to each node. GenBank accession numbers for the sequences used in these analyses are provided in Supplementary Information S1.
doi:10.1371/journal.pone.0003541.g006

and SOPE *sogR1* and *sogR2* sequences were placed in separate subclades with 100% and 98% bootstrap support, respectively. Thus, it seems likely that these two genes in *S. glossinidius* and SOPE have independent phylogenetic origins and did not arise from a recent gene duplication event.

Estimating *d*N/*d*S (the number of nonsynonymous base substitutions per nonsynonymous site versus the number of synonymous base substitutions per synonymous site) provides a means to assess the strength and direction of natural selection acting upon genic (coding) sequences. In bacteria, pairwise estimates of *d*N/*d*S typically fall within the range of 0.04 0.2 for functional genes evolving under stabilizing selection [44]. Conversely, genic sequences that have been rendered inactive (pseudogenes) are expected to have *d*N/*d*S ratios that approach parity. Since we made no direct assessment of the functionality of the truncated *carA* homologs in *S. glossinidius* and SOPE, or the *lux* gene homologs in SOPE, we reasoned that it would be useful to perform tests of selection on these putative coding sequences. Thus, we obtained pairwise estimates of *d*N/*d*S for the *luxI*, *luxR* and *carA* sequences from *S. glossinidius*, SOPE and close free-living relatives (Table 1). Notably, the estimates of *d*N/*d*S in pairwise sequence comparisons between *S. glossinidius* and SOPE for *sogI* (0.071), *sogR1* (0.061), *sogR2* (0.116) and the truncated *carA* homolog (0.095) are all within the range expected for genes evolving under stabilizing selection. This suggests that *sogI*, *sogR1*, *sogR2* and the truncated *carA* all remain functional in *S. glossinidius* and SOPE.

Discussion

Sodalis glossinidius, an endosymbiotic bacterium found in tsetse flies (*Glossina* spp.), uses an AHL-based quorum sensing system to modulate gene expression according to cell density. Since *S. glossinidius* has no known lifestyle component outside of the insect host, the quorum sensing system must play an important role in the regulation of bacterial gene expression during symbiosis. The *S. glossinidius* quorum sensing system utilizes at least two distinct regulatory proteins, designated SogI and SogR1. SogI is an AHL synthase that is responsible for the synthesis of a signaling molecule, OHHL. SogR1 is a DNA-binding transcriptional regulator that interacts with OHHL and modulates the expression of target genes in accordance with bacterial population density. *Sodalis glossinidius* also maintains a second LuxR homolog (SogR2) but the function of this protein was not elucidated in this study.

Table 1. *d*N:*d*S ratios computed from pairwise comparisons of genes involved in quorum sensing in *S. glossinidius*, SOPE and related free-living bacteria.

Sequence Pairs	*d*N	*d*S	*d*N/*d*S Ratio
S. glossinidius carA- SOPE *carA*	0.058	0.610	0.095
S. glossinidius carA- *E. carotovora carA*	0.698	1.504	0.464
SOPE *carA*- *E. carotovora carA*	0.694	1.766	0.393
S. glossinidius sogI- SOPE *sogI*	0.036	0.506	0.071
S. glossinidius sogI- *Y. pestis ypeI*	0.466	1.800	0.259
SOPE *sogI*- *Y. pestis ypeI*	n/d*	n/d*	n/d*
S. glossinidius sogR1- SOPE *sogR1*	0.039	0.635	0.061
S. glossinidius sogR1- *Y. pestis ypeR*	0.533	1.484	0.359
SOPE *sogR1*- *Y. pestis ypeR*	n/d*	n/d*	n/d*
S. glossinidius sogR2- SOPE *sogR2*	0.045	0.388	0.116
S. glossinidius sogR2- *E. carotovora expR2*	n/d*	n/d*	n/d*
S. glossinidius sogR2- *E. carotovora carR*	n/d*	n/d*	n/d*
SOPE *sogR2*- *E. carotovora expR2*	0.497	1.970	0.252
SOPE *sogR2*- *E. carotovora carR*	n/d*	n/d*	n/d*
E. carotovora expR2- *E. carotovora carR*	n/d*	n/d*	n/d*

*n/d – not determined due to saturation (*d*S>2) at synonymous sites. GenBank accession numbers for the sequences used in these analyses are provided in Supplementary Information S1.

doi:10.1371/journal.pone.0003541.t001

In the current study, we used a tiling microarray to identify genes in the *S. glossinidius* genome whose expression is modulated in the presence of the quorum sensing signaling molecule, OHHL. The most striking finding was that OHHL increased the expression of a large number of genes that are known to be involved in the cellular response to oxidative stress. These includes *oxyR*, the positive regulator of hydrogen peroxide inducible genes [45], and numerous genes whose protein products are known to be involved in the direct detoxification of ROS [46,47], including catalase (*katA*), manganese superoxide dismutase (*sodA*), peroxidase (*ahpC*) and glutathione S-transferase (*yfcG*). We also detected increased expression of genes encoding proteins known to be involved in the repair of cellular components following ROS-mediated damage [48 54]. These include thioredoxin reductase (*trxB*), methionine sulfoxide reductase B (*msrB*) and several chaperones known to be involved in protein refolding and maturation under oxidative stress (e.g., *clpB*, *dnaK*, *grpE*, *htpG*, *ibpA*, *nfuA*).

The microarray data also shows that OHHL modulated the expression of genes involved in metal ion transport and storage. In many bacteria, the expression of genes involved in iron and manganese transport and metabolism is coordinately regulated with genes involved in the oxidative stress response [55,56]. Both manganese and iron are known to serve as essential co-factors in a number of enzymes that detoxify ROS (e.g., catalase and superoxide dismutase). Also, soluble iron is known to contribute directly to oxidative stress by catalyzing the formation of free radicals through the Fenton reaction [46]. OHHL was also found to increase the expression of genes involved in the production of iron siderophores and bacterioferritin, which have been shown to play a role in ameliorating oxidative stress by removing toxic iron from solution and (in the case of bacterioferritin) serving as iron storage complexes for heme enzymes [57 60]. In *S. glossinidius*, OHHL also increased expression of genes encoding an iron/manganese transport system (*sitABCD*) that is known to be associated with resistance to oxidative stress in other bacteria [61,62,63]. Furthermore, biochemical assays showed that *S. glossinidius* produced and secreted increased amounts of iron siderophores into culture media in the presence of exogenous OHHL. Finally, it should be noted that OHHL increased transcription of a cluster of genes encoding a polyol metabolism and transport system in *S. glossinidius* (SG0608 SG0614; 1.4 3 fold upregulation). This is conspicuous because polyols are known to be abundant in animal cells under conditions of hyperglycemia [64].

Many insect endosymbionts, including *S. glossinidius*, are known to maintain extremely high infection densities within host tissues and cells over prolonged periods of time [65,66,67]. In a recent study, Heddi et al. [68] compared the metabolic and transcriptional profiles of maize weevil bacteriocytes in the presence and absence of a symbiotic bacterium (SZPE) that is another close relative of *S. glossinidius*. Notably, symbiont-infected cells were found to display an unusual profile of carbohydrate transport and metabolism characterized by the induction of the polyol pathway through increased expression of aldose reductase enzymes. This adaptation is predicted to occur in response to the high carbohydrate uptake and intense metabolic activity in densely

infected bacteriocytes, which generates conditions analogous to those found in mammalian diabetic cells. One critical consequence of this intense metabolic activity is the generation of increased concentrations of ROS [64]. Under these conditions, the weevil bacteriocytes were found to display increased expression of genes encoding proteins that are expected to ameliorate the deleterious effects of ROS.

Many bacteria are known to undergo an "adaptive response" upon challenge with a sublethal dose of ROS [69]. The sublethal challenge serves to prime the expression of a large number of genes involved in the oxidative stress response, rendering the cells more resistant to a subsequent lethal dose. Our findings suggest that quorum sensing serves to prime the cellular response to oxidative stress in *S. glossinidius* in a similar way. The use of a quorum sensing system for this purpose makes sense because the concentrations of ROS are expected to increase in accordance with the density of bacterial infection in host cells and tissues. Thus it seems that symbionts and host cells work together to modulate their gene expression profiles and metabolic activities to minimize the deleterious effects of oxidative stress encountered during their symbiotic interactions. This likely represents a key adaptation in the symbiotic relationship because it allows host cells to sustain dense intracellular symbiont infections.

The discovery of complementary responses towards oxidative stress in the weevil bacteriocytes and in *S. glossinidius* prompted us to search for the presence of quorum sensing regulatory genes in the grain weevil symbiont, SOPE. The SOPE genome sequence was indeed found to contain intact homologs of the *S. glossinidius sogI*, *sogR1* and *sogR2* genes. Phylogenetic analyses indicate that the *lux* genes were present in the last common ancestor of *S. glossinidius* and SOPE, so their acquisition likely predates the origin of the symbiotic associations involving these bacteria. Although the functions of the SOPE *lux* genes have not been defined in this study, molecular evolutionary analyses indicate that *sogI*, *sogR1* and *sogR2* are evolving under stabilizing selection, consistent with the maintenance of gene function in the weevil symbiosis. Given that many bacterial symbionts, including *S. glossinidius* and SOPE, are known to attain extremely high infection densities in host tissues in spite of the burden of oxidative stress, it seems likely that such high infection densities confer important advantages within the symbiosis. For this reason, we expect other insect-bacterial associations to have similar adaptations to deal the challenge of oxidative stress. Furthermore, it is noteworthy that insect symbionts such as *S. glossinidius* and members of the candidate genus *Arsenophonus* have proved difficult to isolate and manipulate in pure culture because they demonstrate a high level of sensitivity towards oxidative stress when cultures are maintained on agar plates at low cell densities [70]. Thus, increasing our understanding of the oxidative stress response in symbiosis may also contribute to the development of improved techniques for the culture and manipulation of symbionts.

Following the establishment of symbiosis, the switch to a strictly vertical mode of transmission and the adoption of a static lifestyle leads to a process of genome streamlining in symbiotic bacteria that is characterized by the progressive inactivation and deletion of genes evolving under relaxed selection [18]. This evolutionary trajectory is largely degenerative and irreversible because symbionts have little or no opportunity to replenish their genetic inventory through parasexual recombination. For example, in *S. glossinidius*, only 50% of the whole genome sequence is anticipated to encode functional proteins [43]. The remaining DNA is composed largely of pseudogenes genes with reading frames truncated by more than 50% as a result of frameshifts and/or premature stop codons. In the current study we found that one such "pseudogene," *carA* (formally a component of an antibiotic biosynthesis gene cluster), unexpectedly displayed a 5-fold increase in transcription in response to OHHL. Pairwise analyses of dN/dS ratios show that *carA* is evolving under strong stabilizing selection in *S. glossinidius* and SOPE. Taken together, these observations suggest that *carA* is maintained as a functional protein coding gene, albeit in a truncated form. This illustrates the potential for adaptive neofunctionalization to occur as a component of the process of genome degeneration and streamlining.

Methods

Bacterial Strains and Culture Conditions

A complete list of plasmids and strains used in this study is provided in Supplementary Table S1. *Sodalis glossinidius* was maintained at 25°C in the semi-defined liquid Mitsuhashi-Maramorosch (MM) medium as described previously [65] or in a defined liquid medium containing 0.15 g/L $CaCl_2$, 0.046 g/L $MgCl_2$, 0.2 g/L KCl, 7.0 g/L NaCl, 4.0 g/L glucose and 6 g/L deferrated casamino acids. *Escherichia coli* strains were maintained at 37°C in either Luria-Bertani (LB) medium or in M9 liquid minimal medium salts supplemented with 0.24 mg/L $MgSO_4$, 0.01 mg/L $CaCl_2$, 4 g/L lactose and 2 g/L casamino acids. *Agrobacterium tumefaciens* KYC55 (pJZ372) (pJZ384) (pJZ410) was maintained at 28°C in AT minimal medium [24]. *Chromobacterium violaceum* CV026 and *P. aeruginosa* PAO1 were maintained at 30°C and 37°C, respectively, in LB medium [25]. Where appropriate, antibiotics were added to the media at the following concentrations: 100 μg/ml of ampicillin, 100 μg/ml of gentamycin, 30 μg/ml (chromosomal integrations) or 50 μg/ml (high copy number plasmids) of kanamycin, 100 μg/ml of spectinomycin and 1.5 μg/ml of tetracycline.

Extraction of Culture Supernatants

Acylated homoserine lactones were extracted from culture supernatants as described by Shaw et al. [71]. Briefly, 500 ml of a stationary phase culture of *S. glossinidius* (grown in liquid MM medium), *E. coli* [TOP10 and TOP10 (pSGI) grown in defined liquid medium] and *P. aeruginosa* were pelleted by centrifugation (8,000×g, 20 min., 4°C). The resulting culture supernatants were filtered through 0.2 μm pore-size membrane filters (Nalgene Labware, Cat. No. 154-0020) and extracted twice with an equal volume of ethyl acetate. The extracts were combined, dried with anhydrous magnesium sulfate, filtered, evaporated using a vacuum centrifuge and resuspended in 500 μL of acetonitrile.

Thin Layer Chromatography Overlay Assay

Samples were chromatographed on C_{18} reverse-phase TLC plates (Whatman, Cat. No. 4801-800) using methanol:water (60:40). Following development, plates were dried and overlaid with live cultures of *A. tumefaciens* KYC55 (pJZ372) (pJZ384) (pJZ410) or *C. violaceum* CV026 indicator strains as described previously [24,25].

High Performance Liquid Chromatography and Mass Spectrometry Analysis

S. glossinidius and *E. coli* TOP10 (pSGI) culture supernatant extractions were fractionated using reverse-phase high performance liquid chromatography. The samples were chromatographed on a C_{18} column with a linear water-acetonitrile gradient (0 40%) containing 0.1% trifluoroacetic acid (v/v). The eluted fractions were screened for the presence of AHL with the *A. tumefaciens* biosensor strain [24]. Bioactive fractions from the *E. coli* TOP10 (pSGI) sample were evaporated using a vacuum centrifuge

and resuspended in methanol. The samples were analyzed by electrospray ionization and CID-MS using a Micromass Quattro II - Triple Quadrupole Mass Spectrometer under positive-ion conditions. Fractions were injected into the mass spectrometer in a solvent containing 50% methanol, 49.9% water and 0.1% formic acid at a flow rate of 5 µL/min.

β-Galactosidase Assay

E. coli strain BW25113, harboring specific reporter plasmids, were grown overnight in LB medium. Overnight cultures were diluted 1:500 in fresh medium containing 0, 4 nM, 40 nM, 0.4 µM, 4 µM, 40 µM, 100 µM or 200 µM of N-(3-oxohexanoyl)-homoserine lactone (Sigma Aldrich, Cat. No. K3255). The cultures were maintained for 3 hours at 37°C and β-galactosidase activity was measured as described by Miller [72]. All assays were conducted in triplicates.

Artificial Induction of Quorum and RNA Isolation

A culture of *S. glossinidius* was grown to early log phase ($OD_{600}\approx0.04$, approximately 1.5×10^7 CFU/ml) in MM liquid medium. This culture was separated to yield eight cultures of equal volume and 100 µM of N-(3-oxohexanoyl)-homoserine lactone (Sigma Aldrich) was added to four of the replicates. After 12 hours of incubation at 25°C, the cells were harvested by centrifugation (5,000×g, 10 min., 4°C). RNA was prepared using the SV Total RNA Isolation System (Promega, Cat. No. Z3100), according to the kit instructions. Aliquots of the RNA samples were treated with DNase I (Ambion, Cat. No. AM1907) to remove contaminating DNA. RNA samples were then reverse transcribed using the TaqMan Reverse Transcription Reagents (Applied Biosystems, Part No. N808-0234), according to the manufacturers instructions. Artificial induction of quorum was verified by measuring relative transcript levels of *sogI* in induced versus uninduced samples using qPCR.

Microarray Expression Analyses and Quantitative PCR

A custom tiling microarray was designed using the eArray Design Creation online application (Agilent Technologies). The array consisted of approximately 40,000 sixty-mer oligonucleotide probes, one probe per 200 bp, covering the entire *Sodalis glossinidius* genome (GenBank accession number AP008232.1). Standard positive and negative control features were also included on the array. Microarrays were printed using Agilent SurePrint technology in the 4×44 k slide format.

Microarray hybridization was performed according to the Agilent Technologies protocol, with the following modifications. Briefly, poly(A) tails were added to the 3′ end of the RNA molecules using the Ambion Poly(A) tailing kit (Cat. No. 1350). The polyadenylated RNA was then used as a template to generate fluorescently labeled cRNA using the Agilent Two-Color Low RNA Input Linear Amplification Kit, labeling OHHL+ RNA with cyanine 3-cytosine triphosphate (CTP) and OHHL- RNA with cyanine 5-CTP, and *vice versa*. Fluorescently labeled cRNA samples (825 ng each), in addition to Agilent RNA spike-in controls, were then fragmented and hybridized to the tiling microarray using the Agilent 2-color GE Hybridization/Wash protocol. Hybridized slides were then scanned in an Agilent Technologies G2505B Microarray Scanner at 5 µm resolution, performing a simultaneous detection of Cyanine-3 and Cyanine-5 signal on the hybridized slide. An extended dynamic range scan was then accomplished by performing a primary scan at 100% laser power and a secondary scan at 10% power; the former used to calculate intensities for non-saturating features, and the latter used to calculate intensities for saturating features. The scanned microarray image files were then loaded into Agilent Feature Extraction Software (v. 9.5.1), which was used to perform calculations that included feature intensities, background measurements, and statistical analyses. To control for variation in individual probe hybridization efficiency and fluorescence intensity, mechanically sheared whole genomic DNA was used as a template for cRNA hybridization on a separate array. Differences in individual probe intensities were used to normalize the raw experimental data. Hierarchical clustering analyses correctly broke down normalized data by sample treatment (+/− OHHL, dye type; r^2 values >0.95).

Statistical analysis of the results was performed using Tiling Microarray Analysis Tools 2 (http://sourceforge.net/projects/timat2). CyberT was used to estimate a confidence in the differential expression by calculating the posterior probability of differential expression [73]. Affymetrix's Integrated Genome Browser was used to visualize the microarray analysis data (http://www.affymetrix.com/support/developer/tools/download_igb.affx).

Microarray results were verified by qPCR. Briefly, RNA samples from the four biological replicates were pooled together, and subjected to DNase I treatment (Ambion, Cat. No. AM1907) until no DNA could be detected by qPCR. RNA samples were then reverse transcribed using the TaqMan Reverse Transcription Reagents (Applied Biosystems, Part No. N808-0234), according to the manufacturers instructions. Reactions were performed in triplicates using iQ Supermix (Bio-Rad, Cat. No. 170-8862), and the samples were quantitated using an iCycler iQ Multicolor Real-Time PCR Detection System (Bio-Rad). Relative transcript levels were estimated using the standard curve method described previously by Dale et al. [74]. In order to validate the microarray results, 10 different loci that were either upregulated, downregulated or displayed no changed in expression profile in the microarray experiment were selected for the qPCR verification.

Siderophore Quantitative Assay

A 150 ml culture of *S. glossinidius* was grown to early log phase ($OD_{600}\approx0.04$, approximately 1.5×10^7 CFU/ml) in MM liquid medium. The cells were pelleted by centrifugation (4,500×g, 10 min., 15°C) and washed twice in an equal volume of 0.85% NaCl solution. After a third centrifugation, the cells were resuspended in 15 ml of 0.85% NaCl solution and 1 ml of this cell suspension was used to inoculate six replicates of 30 ml of defined medium. 100 µM N-(3-oxohexanoyl)-homoserine lactone (Sigma Aldrich, Cat. No. K3255) was added to three of the replicates. After 24 and 48 hours of incubation at 25°C, culture aliquots were pelleted by centrifugation (8,000×g, 20 min., 4°C), supernatants were filtered through 0.2 µm pore-size membrane filters (Millipore, Cat. No. SLGP033RB) and siderophore units were estimated using the chrome azurol S liquid assay [75]. To control for the effects of N-(3-oxohexanoyl)-homoserine lactone as a potential iron chelating agent [76], 100 µM of N-(3-oxohexanoyl)-homoserine lactone was added to all supernatants and reference readings prior to quantitation.

Phylogenetic and Molecular Evolutionary Analyses

Phylogenetic analyses were performed on homologs of *luxI* and *luxR* from *S. glossinidius*, a closely related grain weevil endosymbiont (SOPE) [67], and other closely related Gram negative bacteria known to maintain AHL-based quorum sensing systems. The nucleotide sequences of *luxI* and *luxR* were aligned according to the corresponding protein sequence alignments generated in CLUSTAL, to provide in-frame nucleotide sequence alignments. All third codon position characters were then excluded from subsequent analyses to improve the signal to noise ratio (third

position characters were deemed uninformative due to high levels of substitution at synonymous sites). Maximum likelihood (ML) analyses were performed in PAUP 4.0 [77], using the heuristic tree-bisection-reconnection algorithm. ML parameters were estimated from an initial neighbor-joining tree and optimized in the construction of ML trees using variable base frequencies, a symmetrical substitution matrix and gamma distributed rate variation among sites. Bootstrap values were obtained from analysis of 100 replicates.

Pairwise molecular evolutionary sequence analyses were performed in MEGA 4.0 [78]. Nucleotide sequence alignments were generated in frame, as described above. The frequencies of synonymous and nonsynonymous substitutions (dS and dN, respectively) were estimated using the Kumar method. Z-tests of selection were then used to estimate the probability of rejecting a null hypothesis of strict neutrality in favor of an alternative hypothesis of stabilizing (purifying) selection.

Supporting Information

Table S1 Plasmids and Strains Used in this Study
Found at: doi:10.1371/journal.pone.0003541.s001 (0.06 MB DOC)

Table S2 Results from Microarray Expression Analyses
Found at: doi:10.1371/journal.pone.0003541.s002 (2.68 MB DOC)

Supplementary Information S1
Found at: doi:10.1371/journal.pone.0003541.s003 (0.03 MB DOC)

Acknowledgments

The authors thank Serap Aksoy (Yale University) for providing a culture of *S. glossinidius*, Sang-Jin Suh (Auburn University) for providing a culture of *P. aeruginosa* PAO1, Stephen Winans (Cornell University) for providing the *A. tumefaciens* reporter strain, Robert Simons (University of California, Los Angeles) for providing the plasmid pRS415 and Elliot Rachlin (University of Utah) for assistance with the mass spectrometry.

Author Contributions

Conceived and designed the experiments: MHP CD. Performed the experiments: MHP MB KS. Analyzed the data: MHP RL KO CD. Contributed reagents/materials/analysis tools: CD. Wrote the paper: MHP CD.

References

1. Perret X, Staehelin C, Broughton WJ (2000) Molecular basis of symbiotic promiscuity. Microbiol Mol Biol Rev 64: 180 201.
2. Broughton WJ, Hanin M, Relic B, Kopcinska J, Golinowski W, et al. (2006) Flavonoid-inducible modifications to rhamnan O antigens are necessary for *Rhizobium* sp. strain NGR234-legume symbioses. J Bacteriol 188: 3654 3663.
3. Campbell GR, Reuhs BL, Walker GC (2002) Chronic intracellular infection of alfalfa nodules by *Sinorhizobium meliloti* requires correct lipopolysaccharide core. Proc Natl Acad Sci U S A 99: 3938 3943.
4. Fraysse N, Jabbouri S, Treilhou M, Couderc F, Poinsot V (2002) Symbiotic conditions induce structural modifications of *Sinorhizobium* sp. NGR234 surface polysaccharides. Glycobiology 12: 741 748.
5. Niehaus K, Lagares A, Puhler A (1998) A *Sinorhizobium meliloti* lipopolysaccharide mutant induces effective nodules on the host plant *Medicago sativa* (alfalfa) but fails to establish a symbiosis with *Medicago truncatula*. Mol Plant Microbe Interact 11: 906 914.
6. Dazzo FB, Truchet GL, Hollingsworth RI, Hrabak EM, Pankratz HS, et al. (1991) *Rhizobium* lipopolysaccharide modulates infection thread development in white clover root hairs. J Bacteriol 173: 5371 5384.
7. Skorupska A, Janczarek M, Marczak M, Mazur A, Krol J (2006) Rhizobial exopolysaccharides: genetic control and symbiotic functions. Microb Cell Fact 5.
8. Quandt J, Hillemann A, Niehaus K, Arnold W, Puhler A (1992) An osmorevertant of *Rhizobium meliloti ndvB* deletion mutant forms infection threads but is defective in bacteroid development. Mol Plant Microbe Interact 5: 420 427.
9. Nyholm SV, Deplancke B, Gaskins HR, Apicella MA, McFall-Ngai MJ (2002) Roles of *Vibrio fischeri* and nonsymbiotic bacteria in the dynamics of mucus secretion during symbiont colonization of *the Euprymna scolopes* light organ. Appl Environ Microbiol 68: 5113 5122.
10. McFall-Ngai M, Brennan C, Weis V, Lamarcq L (1998) Mannose adhesin-glycan interactions in the *Euprymna scolopes-Vibrio fischeri* symbiosis. In: Le Gal Y, Halvorson HO, eds. New developments in marine biotechnology. New York: Plenum Press. pp 273 276.
11. Yip ES, Geszvain K, DeLoney-Marino CR, Visick KL (2006) The symbiosis regulator RscS controls the *syp* gene locus, biofilm formation and symbiotic aggregation by *Vibrio fischeri*. Mol Microbiol 62: 1586 1600.
12. DeLoney-Marino CR, Wolfe AJ, Visick KL (2003) Chemoattraction of *Vibrio fischeri* to serine, nucleosides, and N-acetylneuraminic acid, a component of squid light-organ mucus. Appl Environ Microbiol 69: 7527 7530.
13. Kaplan HB, Greenberg EP (1985) Diffusion of autoinducer is involved in regulation of the *Vibrio fischeri* luminescence system. J Bacteriol 163: 1210 1214.
14. Visick KL, Foster J, Doino J, McFall-Ngai M, Ruby EG (2000) *Vibrio fischeri lux* genes play an important role in colonization and development of the host light organ. J Bacteriol 182: 4578 4586.
15. Koropatnick TA, Engle JT, Apicella MA, Stabb EV, Goldman WE, et al. (2004) Microbial factor-mediated development in a host-bacterial mutualism. Science 306: 1186 1188.
16. Wernegreen JJ (2002) Genome evolution in bacterial endosymbionts of insects. Nat Rev Genet 3: 850 861.
17. Zientz E, Dandekar T, Gross R (2004) Metabolic interdependence of obligate intracellular bacteria and their insect hosts. Microbiol Mol Biol Rev 68: 745 777.
18. Dale C, Moran NA (2006) Molecular interactions between bacterial symbionts and their hosts. Cell 126: 453 465.
19. Baumann P (2005) Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu Rev Microbiol 59: 155 189.
20. Aksoy S, Chen X, Hypsa V (1997) Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera:Glossinidae). Insect Mol Biol 6: 183 190.
21. Moran NA, Russell JA, Koga R, Fukatsu T (2005) Evolutionary relationships of three new species of *Enterobacteriaceae* living as symbionts of aphids and other insects. Appl Environ Microbiol 71: 3302 3310.
22. Tsuchida T, Koga R, Meng XY, Matsumoto T, Fukatsu T (2005) Characterization of a facultative endosymbiotic bacterium of the pea aphid *Acyrthosiphon pisum*. Microb Ecol 49: 126 133.
23. Gonzalez JE, Keshavan ND (2006) Messing with Bacterial Quorum Sensing. Microbiol Mol Biol Rev 70: 859 875.
24. Zhu J, Chai Y, Zhong Z, Li S, Winans SC (2003) *Agrobacterium* bioassay strain for ultrasensitive detection of N-acylhomoserine lactone-type quorum-sensing molecules: detection of autoinducers in *Mesorhizobium huakuii*. Appl Environ Microbiol 69: 6949 6953.
25. McClean KH, Winson MK, Fish L, Taylor A, Chhabra SR, et al. (1997) Quorum sensing in *Chromobacterium violaceum*: exploitation of violacein production and inhibition for the detection of N-acylhomoserine lactones. Microbiology 143: 3703 3711.
26. Morin D, Grasland B, Vallee-Rehel K, Dufau C, Haras D (2003) On-line high-performance liquid chromatography-mass spectrometric detection and quantification of N-acylhomoserine lactones, quorum sensing signal molecules, in the presence of biological matrices. J Chromatogr A 1002: 79 92.
27. Winson MK, Camara M, Latifi A, Foglino M, Chhabra SR, et al. (1995) Multiple N-acyl-L-homoserine lactone signal molecules regulate production of virulence determinants and secondary metabolites in *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A 92: 9427 9431.
28. Egland KA, Greenberg EP (1999) Quorum sensing in *Vibrio fischeri*: elements of the *luxI* promoter. Mol Microbiol 31: 1197 1204.
29. Fuqua WC, Winans SC (1994) A LuxR-LuxI type regulatory system activates *Agrobacterium* Ti plasmid conjugal transfer in the presence of a plant tumor metabolite. J Bacteriol 176: 2796 2806.
30. Lewenza S, Sokol PA (2001) Regulation of ornibactin biosynthesis and N-acyl-L-homoserine lactone production by CepR in *Burkholderia cepacia*. J Bacteriol 183: 2212 2218.
31. Fuqua C, Winans SC, Greenberg EP (1996) Census and consensus in bacterial ecosystems: the LuxR-LuxI family of quorum-sensing transcriptional regulators. Annu Rev Microbiol 50: 727 751.
32. Darby AC, Lagnel J, Matthew CZ, Bourtzis K, Maudlin I, et al. (2005) Extrachromosomal DNA of the symbiont *Sodalis glossinidius*. J Bacteriol 187: 5003 5007.
33. Rolfe B, Eisenberg MA (1968) Genetic and biochemical analysis of the biotin loci of *Escherichia coli* K-12. J Bacteriol 1968 96: 515 524.
34. Derzelle S, Duchaud E, Kunst F, Danchin A, Bertin P (2002) Identification, characterization, and regulation of a cluster of genes involved in carbapenem biosynthesis in *Photorhabdus luminescens*. Appl Environ Microbiol 68: 3780 3789.

35. McGowan SJ, Sebaihia M, Porter LE, Stewart GSAB, Williams P, et al. (1996) Analysis of bacterial carbapenem antibiotic production genes reveals a novel beta-lactam biosynthesis pathway. Mol Microbiol 22: 415 426.
36. McGowan SJ, Sebaihia M, O'Leary S, Hardie KR, Williams P, et al. (1997) Analysis of the carbapenem gene cluster of *Erwinia carotovora*: definition of the antibiotic biosynthetic genes and evidence for a novel beta-lactam resistance mechanism. Mol Microbiol 26: 545 556.
37. Thomson NR, Crow MA, McGowan SJ, Cox A, Salmond GPC (2000) Biosynthesis of carbapenem antibiotic and prodigiosin pigment in *Serratia* is under quorum sensing control. Mol Microbiol 36: 539 556.
38. Bainton NJ, Stead P, Chhabra SR, Bycroft BW, Salmond GP, et al. (1992) *N*-(3-oxohexanoyl)-L-homoserine lactone regulates carbapenem antibiotic production in *Erwinia carotovora*. Biochem J 288: 997 1004.
39. Andersson RA, Eriksson ARB, Heikinheimo R, Mäe A, Pirhonen M, et al. (2000) Quorum sensing in the plant pathogen *Erwinia carotovora* subsp. *carotovora*: the role of *expR* (*Ecc*). Mol Plant Microbe Interact 13: 384 393.
40. McGowan S, Sebaihia M, Jones S, Yu B, Bainton N, et al. (1995) Carbapenem antibiotic production in *Erwinia carotovora* regulated by CarR, a homologue to the LuxR transcriptional activator. Microbiology 141: 541 550.
41. Heddii A, Charles H, Khatchadourian C, Bonnot G, Nardon P (1998) Molecular characterization of the principal symbiotic bacteria of the weevil *Sitophilus oryzae*: a peculiar G+C content of an endocytobiotic DNA. J Mol Evol 47: 52 61.
42. Dale C, Plague GR, Wang B, Ochman H, Moran NA (2002) Type III secretion systems and the evolution of mutualistic endosymbiosis. Proc Natl Acad Sci U S A 99: 12397 12402.
43. Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, et al. (2006) Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Res 16: 149 156.
44. Rocha EP, Smith JM, Hurst LD, Holden MT, Cooper JE, et al. (2006) Comparisons of dN/dS are time dependent for closely related bacterial genomes. J Theor Biol 239: 226 235.
45. Bauer CE, Elsen S, Bird TH (1999) Mechanisms for redox control of gene expression. Annu Rev Microbiol 53: 495 523.
46. Imlay JA (2003) Pathways of oxidative damage. Annu Rev Microbiol 57: 395 418.
47. Kanai T, Takahashi K, Inoue H (2006) Three distinct-type glutathione S-transferases from *Escherichia coli* important for defense against oxidative stress. J Biochem 140: 703 711.
48. Angelini S, Gerez C, Choudens SO-d, Sanakis Y, Fontecave M, et al. (2008) NfuA, a new factor required for maturing Fe/S proteins in *Escherichia coli* under oxidative stress and iron starvation conditions. J Biol Chem 283: 1408 14091.
49. Hossain MM, Nakamoto H (2003) Role for the cyanobacterial HtpG in protection from oxidative stress. Curr Microbiol 46: 70 76.
50. Kitagawa M, Miyakawa M, Matsumura Y, Tsuchido T (2002) *Escherichia coli* small heat shock proteins, IbpA and IbpB, protect enzymes from inactivation by heat and oxidants. Eur J Biochem 269: 2907 2917.
51. Matuszewska E, Kwiatkowska J, Kuczynska-Wisnik D, Laskowska E (2008) *Escherichia coli* heat-shock proteins IbpA/B are involved in resistance to oxidative stress induced by copper. Microbiology 154: 1739 1747.
52. Stewart EJ, Aslund F, Beckwith J (1998) Disulfide bond formation in the *Escherichia coli* cytoplasm: an in vivo role reversal for the thioredoxins. EMBO J 17: 5543 5550.
53. Takemoto T, Zhang QM, Yonei S (1998) Different mechanisms of thioredoxin in its reduced and oxidized forms in defense against hydrogen peroxide in *Escherichia coli*. Free Radical Biol & Med 24: 556 562.
54. Weissbach H, Resnick L, Brot N (2005) Methionine sulfoxide reductases: history and cellular role in protecting against oxidative damage. Biochim Biophys Acta 1703: 203 212.
55. Hantke K (2001) Iron and metal regulation in bacteria. Curr Opin Microbiol 4: 172 177.
56. Touati D (2000) Iron and oxidative stress in bacteria. Arch Biochem Biophys 373: 1 6.
57. Carrondo MA (2003) Ferritins, iron uptake and storage from the bacterioferritin viewpoint. EMBO J 22: 1959 1968.
58. Chen CY, Morse SA (1999) *Neisseria gonorrhoeae* bacterioferritin: structural heterogeneity, involvement in iron storage and protection against oxidative stress. Microbiology 145: 2967 2975.
59. Ma J-F, Ochsner UA, Klotz MG, Nanayakkara VK, Howell ML, et al. (1999) Bacterioferritin A modulates catalase A (KatA) activity and resistance to hydrogen peroxide in *Pseudomonas aeruginosa*. J Bacteriol 181: 3730 3742.
60. Tindale AE, Mehrotra M, Ottem D, Page WJ (2000) Dual regulation of catecholate siderophore biosynthesis in *Azotobacter vinelandii* by iron and oxidative stress. Microbiology 146: 1617 1626.
61. Davies BW, Walker GC (2007) Disruption of *sitA* compromises *Sinorhizobium meliloti* for manganese uptake required for protection against oxidative stress. J Bacteriol 189: 2101 2109.
62. Kitphati W, Ngok-ngam P, Suwanmaneerat S, Sukchawalit R, Mongkolsuk S (2007) *Agrobacterium tumefaciens fur* has important physiological roles in iron and manganese homeostasis, the oxidative stress response, and full virulence. Appl Envir Microbiol 73: 4760 4768.
63. Sabri M, Léveillé S, Dozois CM (2006) A SitABCD homologue from an avian pathogenic *Escherichia coli* strain mediates transport of iron and manganese and resistance to hydrogen peroxide. Microbiology 152: 745 758.
64. Brownlee M (2001) Biochemistry and molecular cell biology of diabetic complications. Nature 414: 813 820.
65. Dale C, Maudlin I (1999) *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. Int J Syst Bacteriol 49: 267 275.
66. Mira A, Moran NA (2002) Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. Microb Ecol 44: 137 143.
67. Heddi A, Grenier AM, Khatchadourian C, Charles H, Nardon P (1999) Four intracellular genomes direct weevil biology: nuclear, mitochondrial, principal endosymbiont, and *Wolbachia*. Proc Natl Acad Sci U S A 96: 6814 6819.
68. Heddi A, Vallier A, Anselme C, Xin H, Rahbe Y, et al. (2005) Molecular and cellular profiles of insect bacteriocytes: mutualism and harm at the initial evolutionary step of symbiogenesis. Cell Microbiol 7: 293 305.
69. Crawford DR, Davies KJA (1994) Adaptive response and oxidative stress. Environ Health Perspect 102(Suppl. 10): 25 28.
70. Pontes MH, Dale C (2006) Culture and manipulation of insect facultative symbionts. Trends Microbiol 14: 406 412.
71. Shaw PD, Ping G, Daly SL, Cha C, Cronan JE, et al. (1997) Detecting and characterizing N-acyl-homoserine lactone signal molecules by thin-layer chromatography. Proc Natl Acad Sci U S A 94: 6036 6041.
72. Miller JH (1992) A short course in bacterial genetics: a laboratory manual and handbook for *Escherichia coli* and related bacteria. New York: Cold Spring Harbor Laboratory Press. pp 72 74.
73. Baldi P, Long AD (2001) A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. Bioinformatics 17: 509 519.
74. Dale C, Jones T, Pontes M (2005) Degenerative evolution and functional diversification of type-III secretion systems in the insect endosymbiont *Sodalis glossinidius*. Mol Biol Evol 22: 758 766.
75. Payne SM (1994) Detection, isolation, and characterization of siderophores. Methods Enzymol 235: 329 344.
76. Kaufmann GF, Sartorio R, Lee SH, Rogers CJ, Meijler MM, et al. (2005) Revisiting quorum sensing: discovery of additional chemical and biological functions for 3-oxo-N-acylhomoserine lactones. Proc Natl Acad Sci U S A 102: 309 314.
77. Swofford DL (2000) PAUP*: Phylogenetic Analysis Using Parsimony and other methods. Sunderland, MA: Sinauer, 4b10 ed:.
78. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596 1599.
79. Simons RW, Houman F, Kleckner N (1987) Improved single and multicopy lac-based cloning vectors for protein and operon fusions. Gene 53: 85 96.
80. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97: 6640 6645.

APPENDIX C

PHYLOGENETIC ANALYSIS OF SYMBIONTS IN FEATHER FEEDING LICE OF THE GENUS COLUMBICOLA: EVIDENCE FOR REPEATED SYMBIONT REPLACEMENTS

Wendy A. Smith, Kevin P. Johnson, Kelly F. Oakeson, David L. Reed, Tamar Carter, Kari L. Smith, Ryuichi Koga, Takema Fukatsu, Dale H. Clayton, and Colin Dale

Abstract

Many groups of insects have obligate bacterial symbionts that are vertically transmitted. Such symbionts typically involve a monophyletic group of bacteria living in a well-defined host clade. These bacteria often have a phylogeny that is congruent with that of the insect host. Here we show that bacteria living in a single genus of feather lice, *Columbicola* (Insecta: Phthiraptera), are an exception to this pattern. The phylogeny of bacteria within *Columbicola* has a star-like topology with very short internodes and long terminal branches. It includes three polyphyletic lineages of *Gammaproteobacteria*, including a distinct symbiont lineage distantly allied to the aphid obligate endosymbiont *Buchnera aphidicola*, a small symbiont clade clustering with *Salmonella enterica*, and a large and well-defined symbiont clade that contains the tsetse facultative symbiont *Sodalis glossinidius* and allied symbionts from diverse insects. Two of these three symbiont lineages exhibit high levels of nucleotide substitution, suggesting accelerated molecular evolution due to relaxed purifying selection, which is typical of many vertically-transmitted insect symbionts. Representatives of the fast-evolving and slow-evolving symbiont lineages exhibit the same localization, migration and transmission patterns in their hosts. This fact is consistent with the hypothesis that repeated symbiont replacements have occurred during the radiation of *Columbicola* spp. Based on these results, we suggest that lice and other insects may experience repeated symbiont replacements through acquisition of free-living progenitors residing in the local environment.

Introduction

Many insects have obligate, primary endosymbiotic bacteria that provide nutrients which are lacking in their diets. Associations between primary symbionts and their insect hosts are often ancient in origin, and have facilitated the exploitation of new ecological niches by insects (Douglas, 1989). The vertical transmission of primary symbionts often leads to repeated bouts of co-speciation, resulting in topological congruence between the insect and symbiont phylogenies (Moran et al., 1993; Baumann, 2005).

Feather lice (Insecta: Phthiraptera) are obligate, permanent ectoparasites of birds and mammals that spend their entire life cycle on the host (Johnson and Clayton, 2003). The genus *Columbicola* contains 88 described morpho-species, all of which parasitize columbiform birds (pigeons and doves) (Bush et al., 2009). Some of these morpho-species are further divided into molecularly distinct cryptic species (Malenke et al., 2009). Species of *Columbicola* are relatively host-specific, with most known from only a single species of bird host. Transmission of lice between birds occurs mainly during periods of direct contact, as occurs between parent birds and their offspring in the nest (Harbison et al., 2008). However, *Columbicola* are also known to disperse phoretically on hippoboscid flies, which are winged parasites of birds (Harbison et al., 2008; Harbison and Clayton, 2011). True to their name, feather lice feed primarily on feathers, as well as dead skin and other dermal "debris" (Clayton et al., 2008). Feathers are a nutritionally challenging diet consisting mostly of hard proteins called keratins, which are difficult to digest, and which have amino acid compositions that are markedly biased (Gillespie and Frenkel, 1974). Moreover, the availability of vitamins and co-factors may be limited in feather diets (Waterhouse, 1957).

A bacterial endocellular symbiont was histologically observed in the abdomen of *Columbicola columbae* decades ago (Ries, 1931); however, molecular aspects of this symbiosis were investigated only recently. Sequencing and phylogenetic analysis of the bacterial 16S rRNA gene and the protein coding gene *fusA* revealed that the symbiont of *C. columbae* is closely related to the facultative symbiont of tsetse flies *Sodalis glossinidius* and allied insect symbionts in the *Gammaproteobacteria*. *In situ* hybridization experiments demonstrated that the symbiont is localized within specialized bacteriocytes and passed to offspring via ovarial transmission (Fukatsu et al., 2007). The function of the *C. columbae* symbiosis is unknown, but it seems likely that the symbiosis has a nutritional basis because keratin-rich feather constitutes a nutritionally incomplete diet (Gillespie and Frenkel, 1974; Marshall et al., 1991).

We investigated the bacterial symbionts of more than 40 members of the genus *Columbicola* sampled from pigeons and doves around the world. Since *Columbicola* symbionts are endocellular in bacteriocytes, and vertically transmitted via ovarial passage, we expected them to exhibit a pattern of co-speciation with their insect hosts. However, in contrast, our molecular phylogenetic analyses revealed striking diversity and evolutionary dynamics in the host-symbiont associations of this single insect genus. We propose and test several hypotheses to account for these unexpected evolutionary patterns.

Materials and Methods

Sample collection and DNA sequencing

Samples of lice were collected from wild birds using the post mortem ruffling procedure described by Clayton and Drown (2001). DNA was extracted from individual

louse samples by first removing the head and placing both the head and abdomen in extraction buffer ATL (Qiagen). DNA was then isolated using a Qiagen DNAeasy Tissue Extraction Kit. Prior to DNA extraction, body parts were removed and mounted in balsam on microscope slides as morphological vouchers.

DNA extracts were used as template for PCR amplification of a segment of the bacterial 16S rRNA gene (1.46-kb). For a subset of our samples we also amplified a 0.76-kb fragment of the elongation factor EF-G (*fusA*) gene and a 1.49kb fragment of the heat shock chaperone (*groEL*) gene. We used the universal primers 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-TACGGTTACCTTGTTACGACTT-3') to amplify 16S rRNA gene, the degenerate primers GroELF1 (5' – ATGGGCWGCWAAAGAYGTRAAAT – 3') and GroELR1 (5' – TCGGTRGTGATMATCAGRCCRGC-3') (designed from an alignment of insect symbionts and other free living members of the *Gammaproteobacteria*) to amplify the *groEL* gene, and FusAF (5'-CAT CGG CAT CAT GGC NCA YAT HGA-3') and FusAR (5'-CAG CAT CGG CTG CAY NCC YTT RTT-3') (Dale et al., 2002) to amplify the *fusA* gene.

The PCR products were purified using the Qiagen gel extraction kit and concentrated in Microcon columns (Millipore). Purified products were cloned into a TOPO 2.1 vector (Invitrogen). Sanger sequencing reactions were performed on 48 clones derived from *C. columbae* and *C. baculoides* DNA, and a minimum of four clones derived from each of the other DNA samples. Sequences were resolved and checked in the software package Lasergene (DNAStar, Inc. Madison, Wisconsin). All symbiont

sequences were deposited in the DDBJ/EMBL/GenBank nucleotide sequence databases under the accession numbers listed in Table C.1.

Molecular phylogenetic analyses

In order to reconstruct the phylogeny of the *Columbicola* spp. symbionts, we used Muscle (Edgar, 2004) to align sequences of the 16S rRNA gene alone, and a combined dataset comprising 16S rRNA, *fusA* and *groEL* genes. Sequences from other symbionts and free-living bacteria were selected for inclusion on the basis of sequence similarity, using the BLAST search tool (NCBI). We used this approach to ensure that, for each *Columbicola* spp. symbiont sequence, the three most closely related symbiont sequences and the most closely related free-living bacterial sequence from GenBank were represented in our dataset. The remaining free-living taxa were selected to provide appropriate resolution within the family Enterobacteriaceae (Table C.2). *Vibrio cholerae* was selected as an outgroup because it represents a distantly related member of the family Enterobacteriaceae. We then used Modeltest (Posada and Crandall, 1998) and JModelTest (for Bayesian Information Criteria; Posada, 2008) to infer the most appropriate model of sequence evolution (GTR + I + G) for subsequent analyses. Analysis of 16S rRNA gene sequences was performed using the maximum likelihood (ML) approach implemented in PhyML (Guindon and Gascuel, 2003), with 25 random starting trees and 100 bootstrap replicates. Bayesian posterior probabilities were estimated using MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001). Runs were carried out for 1.5 million generations using the default parameters of 4 chains (3 heated and one cold) and stopped when the standard deviation of split frequencies converged to less than 0.00001. The first 4000 generations were discarded as burn-in based on the stabilization

Table C.1. Specimens of *Columbicola* used in this study.

#	Louse	Host	Location	Louse Voucher Code	Gene
1	*C. adamsi*	*Patagioenas plumbea*	Guyana	4.24.99.3	16S rDNA
2	*C. arnoldi*	*Macropygia nigrirostris*	Papua New Guinea	5.14.03.5	16S rDNA
					GroEL
3	*C. bacillus*	*Streptopelia decaocto*	Netherlands	11.15.99.1	16S rDNA
4	*C. baculoides*	*Zenaida macroura*	USA	10.19.98.1	16S rDNA
5	*C. claviformis*	*Columba palumbus*	UK	1.20.03.16	16S rDNA
					FusA
6	*C. clayae*	*Treron waalia*	Ghana	3.21.00.9	*FusA*
					GroEL
7	*C. claytoni*	*Ducula rufigaster*	Papua New Guinea	7.26.04.6	16S rDNA
8	*C. claytoni*	*Ducula rufigaster*	Papua New Guinea	8.19.03.14	16S rDNA
					FusA
9	*C. columbae*	*Columba livia*	USA	6.29.98.3	16S rDNA
					FusA
					GroEL
10	*C. columbae*	*Columba livia*	Australia	7.15.02	16S rDNA
11	*C. elbeli*	*Treron sieboldii*	China	6.6.05.4	16S rDNA
					GroEL
12	*C. exilicornis* 1	*Macropygia amboinensis*	Papua New Guinea	8.19.03.7	16S rDNA
					FusA
					GroEL

Table C.1 cont.

#	Louse	Host	Location	Louse Voucher Code	Gene
13	*C. exilicornis* 1	*Macropygia amboinensis*	Papua New Guinea	8.19.03.8	16S rDNA
14	*C. exilicornis* 2	*Phapitreron amethystinus*	Phillipines	5.26.99.6	16S rDNA
15	*C. extinctus*	*Patagioenas fasciata*	USA	1.20.039.1	16S rDNA
16	*C. fortis*	*Otidiphaps nobilis*	Papua New Guinea	5.14.03.7	*FusA*
					GroEL
17	*C. gracilicapitis*	*Leptotila jamaicensis*	Mexico	9.29.98.4	16S rDNA
18	*C. guimaraesi* 1	*Chalcophaps indica*	Vanuatu	7.26.04.4	16S rDNA
19	*C. guimaraesi* 2	*Chalcophaps indica*	Australia	7.20.04.12	16S rDNA
20	*C. harbisoni*	*Phaps histrionica*	Australia	5.14.03.9	16S rDNA
21	*C. koopae*	*Geophaps scripta*	Australia	1.8.03.10	16S rDNA
22	*C. macrourae* 1	*Geotrygon montana*	Mexico	9.29.98.1	16S rDNA
					FusA
23	*C. macrourae* 1	*Leptotila plumbeiceps*	Mexico	10.19.98.4	16S rDNA
24	*C. macrourae* 1	*Leptotila verreauxi*	Mexico	10.19.98.2	16S rDNA
25	*C. macrourae* 2	*Zenaida asiatica*	USA	9.29.98.5	16S rDNA
26	*C. macrourae* 3	*Zenaida macroura*	USA	2.1.99.9	16S rDNA
					GroEL
27	*C. macrourae* 4	*Zenaida galapagoensis*	Galapagos	7.1.99.2	16S rDNA
28	*C. malenkeae*	*Ducula pacifica*	Vanuatu	1.27.04.2	16S rDNA
29	*C. masoni*	*Petrophassa albipennis*	Australia	5.14.03.13	16S rDNA
30	*C. masoni*	*Petrophassa rufipennis*	Australia	1.27.04.12	16S rDNA

Table C.1 cont.

#	Louse	Host	Location	Louse Voucher Code	Gene
					FusA
31	*C. mckeani*	*Ocyphaps lophotes*	Australia	1.20.03.10	16S rDNA
					FusA
					GroEL
32	*C. mjoebergi*	*Geopelia placida*	Australia	5.14.03.17	16S rDNA
					FusA
33	*C. mjoebergi*	*Geopelia striata*	Hawaii	1.20.03.13	16S rDNA
34	*C. mjoebergi*	*Geopelia striata*	Hawaii	3.21.00.5	16S rDNA
35	*C. passerinae* 2	*Claravis pretiosa*	Mexico	9.29.98.3	16S rDNA
36	*C. passerinae* 2	*Claravis pretiosa*	Mexico	2.1.99.6	16S rDNA
37	*C. paradoxus*	*Lopholaimus antarcticus*	Australia	1.27.04.5	16S rDNA
					FusA
38	*C. rodmani*	*Geopelia humeralis*	Australia	5.14.03.12	16S rDNA
39	*C. timmermanni*	*Leptotila rufaxilla*	Guyana	1.08.03.7	16S rDNA
40	*C. timmermanni*	*Leptotila rufaxilla*	Guyana	4.24.99.2	16S rDNA
					FusA
41	*C. tschulyschman*	*Columba livia*	USA	5.7.09.1	16S rDNA
42	*C. veigasimoni*	*Phapitreron leucotis*	Phillipines	5.26.99.3	16S rDNA
43	*C. waggermani*	*Patagioenas leucocephala*	USA	11.15.99.8	16S rDNA
44	*C. waiteae*	*Columba leucomela*	Australia	1.27.04.8	*FusA*
					GroEL

Table C.1 cont.

#	Louse	Host	Location	Louse Voucher Code	Gene
45	*C. waltheri*	*Geotrygon frenata*	Peru	1.20.03.4	16S rDNA
46	*C. wombeyi*	*Geophaps plumifera*	Australia	1.8.03.16	16S rDNA

of log likelihood values at this point. Consensus trees were built based on the 50% majority rule consensus. For the combined analysis, we first used a partition homogeneity test (Farris et al., 1994, 1995; Swofford, 2003) to test for conflict between the three sequences in our combined dataset (16S rRNA, *fusA* and *groEL* genes). Since there were several taxa from which we could only obtain sequence data for two of the three loci, absent sequences were treated as missing data in the tree-building software. For the combined analysis, data were partitioned and parameters were estimated separately for each gene.

Analysis of 16S rRNA secondary structure

In order to determine the impact of substitutions in the *Columbicola* spp. symbiont sequences on the structure of their 16S rRNA molecules, we mapped substitutions from these sequences onto the secondary structure of the 16S rRNA sequence derived from the most closely related outgroup for which a secondary structure has been deduced (*Yersinia pestis*; Cannone et al., 2002). The resulting homology models were then used to compute the relative ratios of substitutions resulting in (i) changes that preserve the secondary structure of the molecule and (ii) changes that induce perturbations in structure (i.e. stem-loop transitions) (Pei et al., 2010). We then further

Table C.2. Other 16S rDNA sequences used in this study.

Isolate	**Accession #**
Buchnera aphidicola str 5A (*Acrythosyphum pisum*)	CP01161
Candidatus *Arsenophonus arthropodicus*	DQ115536
Candidatus *Blochmannia floridanus*	BX248583.1
Candidatus *Blochmannia pennsylvanicus* strain BPEN	CP00016
Endosymbiont of *Haematopinus eurysternus*	DQ076661
Endosymbiont of *Pediculus capitis*	DQ076659
Endosymbiont of *Pediculus humanus*	DQ076660
Endosymbiont of *Polyplax serrata*	DQ076667
Enterobacter endosymbiont of *Metaseiulus occidentalis* clone pAJ240	AY753173
Enterobacter hormaechei subsp. *steigerwaltii* strain EN562T	AJ853890
Dickeya dadantii strain S3-1	AY360397
Escherichia coli K12	U00096
Pantoea agglomerans strain NCTC9381T	AJ251466
Photorhabdus luminescens subsp. *laumondii* TTO1.	BX571859
Primary endosymbiont of *Sitophilus granarius*	AY126638
Primary endosymbiont of *Sitophilus oryzae*	AF548142
Primary endosymbiont of *Sitophilus rugicollis*	AY126639
Primary endosymbiont of *Sitophilus zeamais*	AF548137
Primary symbiont of *Pseudolynchia canariensis*	DQ115535
Salmonella enterica subsp. *enterica* serovar *typhimurium*	AP011957
Secondary endosymbiont of *Bactericera cockerelli*	AF263557
Secondary endosymbiont of *Cantao ocellatus*	AB541010
Secondary endosymbiont of *Craterina melbae* CMS06	EF174495
Secondary endosymbiont of *Curculio sikkimensis*, isolate S86_4	AB541010
Secondary endosymbiont of *Paracoccus nothofagicola*	AF476109
Shigella flexneri 2002017	CP001383
Sodalis glossinidius endosymbiont of tsetse *Glossina morsitans*.	AF548135
Wigglesworthia glossinidia endosymbiont of *Glossina brevipalpis*	BA00021
Yersinia pestis strain KIM	AE009952

analyzed the positions of substitutions in the *Columbicola* spp. symbiont 16S rRNA sequences in accordance with a position-specific variability map computed previously using 3,407 bacterial 16S rRNA sequences (Wuyts et al., 2001). In our analysis, substitutions were scored according to a binary scheme utilized by Pei et al. (2010), in which sites are designated as having substitution rates that are either higher or lower than the average for all sites in the Wuyts et al. (2001) study.

Fluorescent in situ hybridization

Oligonucleotide probes specific to the 16S rRNA sequence from *C. baculoides* (5'-GTTTTCTGTTACCGTTCGATT-3' and 5'-TTGCTTTTTCCTTCTTACTG-3') were used for whole-mount fluorescent *in situ* hybridization as described previously (Fukatsu et al., 2007). Insects obtained from a colony of *C. baculoides* maintained on captive mourning doves (*Zenaida macroura*) were fixed in Carnoy's solution (chloroform-ethanol-acetic acid [6:3:1]) for two days. The lice were then washed three times in ethanol and immersed in 6% (v/v) H_2O_2 in ethanol for 7 d to quench the autofluorescence of insect tissues (Koga et al., 2009). After quenching, the insects were washed three times in ethanol and then decapitated and punctured repeatedly with a fine tungsten needle throughout the abdomen. They were then washed twice in ethanol, three times in phosphate-buffered saline containing 0.3% Triton X-100 (PBSTx), and equilibrated in hybridization buffer (20 mM Tris-HCl [pH 8.0], 0.9 M NaCl, 0.01% sodium dodecyl sulfate, 30% formamide). The probes and SYTOX green were added at final concentrations of 100 nM and 0.5 μM, respectively, and the specimens were incubated overnight at room temperature. The specimens were then washed several times in PBSTx, mounted in Slowfade (Invitrogen) and observed under both an epifluorescence

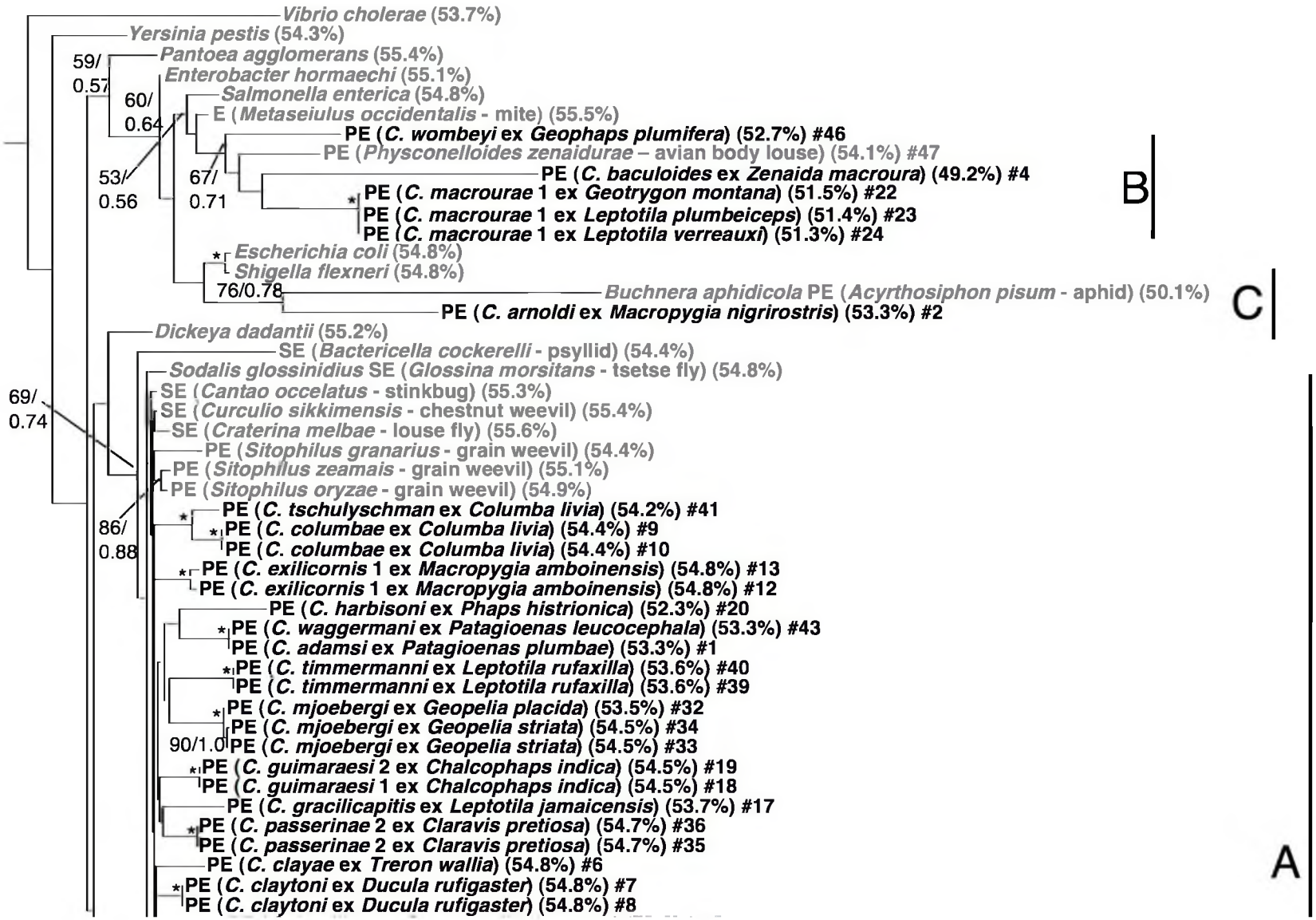
microscope (Axiophoto; Carl Zeiss) and a laser confocal microscope (PASCAL5; Carl Zeiss). To confirm specific detection of the symbionts, a series of control experiments were conducted. These consisted of a no-probe control, RNase digestion control, and a competitive-suppression control with excess unlabeled probe as described previously (Sakurai et al., 2005).

Co-phylogenetic analyses

We used two alternative approaches to test for congruence between louse and symbiont phylogenies. First, we conducted Shimodaira-Hasegawa (S-H) tests (Shimodaira and Hasegawa, 1999) on the best louse tree from Johnson et al., (2007) and the ML trees from the 16S rRNA gene alone (Fig. C.1) and combined sequences (Fig. C.2) from the symbionts. These trees were pruned to include only a single representative sequence from each louse/symbiont taxon. This method was used to assess whether the symbiont data can be used to reject the louse phylogenetic tree.

As a second method, we reconstructed the number of co-speciation events between the symbiont and louse trees using reconciliation analysis (Page, 1990a) as implemented in TreeMap 1 (Page, 1995). This analysis was again performed using both the 16S rRNA gene tree (Fig. C.1) and the combined tree (Fig. C.2). As in the S-H test analysis, both the louse and symbiont trees were pruned to include only a single representative of each louse/symbiont taxon to avoid artificially biasing the results in favor of congruence. The symbiont phylogeny was randomized 10,000 times to determine if the number of inferred co-speciation events was greater than that expected by chance (Page, 1990b, 1995).

Figure C.1. Phylogeny of *Columbicola* spp. symbionts (bold) and related bacteria based on maximum likelihood and Bayesian analyses of a 1.46-kbp fragment of 16S rRNA gene sequences. Insect symbionts are designed by the prefix "PE" (primary endosymbiont), "SE" (secondary endosymbiont) or "E" (if unknown), followed by insect host name and common name (or latin name of bird host for *Columbicola* spp.) The numbers adjacent to nodes indicate maximum likelihood bootstrap values (to left of diagonal line) and Bayesian posterior probabilities (to right of line), for nodes with bootstrap support >50% and Bayesian posterior probabilities >0.5. Asterisks indicate nodes with 100 % bootstrap support and Bayesian posterior probability = 1. Numbers in parentheses represent the G+C content of the 16S rRNA sequences. Final numbers correspond to sample numbers in Table C.1.


Vibrio cholerae (53.7%)
Yersinia pestis (54.3%)
Pantoea agglomerans (55.4%)
Enterobacter hormaechi (55.1%)
Salmonella enterica (54.8%)
E (Metaseiulus occidentalis - mite) (55.5%)
PE (C. wombeyi ex Geophaps plumifera) (52.7%) #46
PE (Physconelloides zenaidurae – avian body louse) (54.1%) #47
PE (C. baculoides ex Zenaida macroura) (49.2%) #4
PE (C. macrourae 1 ex Geotrygon montana) (51.5%) #22
PE (C. macrourae 1 ex Leptotila plumbeiceps) (51.4%) #23
PE (C. macrourae 1 ex Leptotila verreauxi) (51.3%) #24
Escherichia coli (54.8%)
Shigella flexneri (54.8%)
Buchnera aphidicola PE (Acyrthosiphon pisum - aphid) (50.1%)
PE (C. arnoldi ex Macropygia nigrirostris) (53.3%) #2
Dickeya dadantii (55.2%)
SE (Bactericella cockerelli - psyllid) (54.4%)
Sodalis glossinidius SE (Glossina morsitans - tsetse fly) (54.8%)
SE (Cantao occelatus - stinkbug) (55.3%)
SE (Curculio sikkimensis - chestnut weevil) (55.4%)
SE (Craterina melbae - louse fly) (55.6%)
PE (Sitophilus granarius - grain weevil) (54.4%)
PE (Sitophilus zeamais - grain weevil) (55.1%)
PE (Sitophilus oryzae - grain weevil) (54.9%)
PE (C. tschulyschman ex Columba livia) (54.2%) #41
PE (C. columbae ex Columba livia) (54.4%) #9
PE (C. columbae ex Columba livia) (54.4%) #10
PE (C. exilicornis 1 ex Macropygia amboinensis) (54.8%) #13
PE (C. exilicornis 1 ex Macropygia amboinensis) (54.8%) #12
PE (C. harbisoni ex Phaps histrionica) (52.3%) #20
PE (C. waggermani ex Patagioenas leucocephala) (53.3%) #43
PE (C. adamsi ex Patagioenas plumbae) (53.3%) #1
PE (C. timmermanni ex Leptotila rufaxilla) (53.6%) #40
PE (C. timmermanni ex Leptotila rufaxilla) (53.6%) #39
PE (C. mjoebergi ex Geopelia placida) (53.5%) #32
PE (C. mjoebergi ex Geopelia striata) (54.5%) #34
PE (C. mjoebergi ex Geopelia striata) (54.5%) #33
PE (C. guimaraesi 2 ex Chalcophaps indica) (54.5%) #19
PE (C. guimaraesi 1 ex Chalcophaps indica) (54.5%) #18
PE (C. gracilicapitis ex Leptotila jamaicensis) (53.7%) #17
PE (C. passerinae 2 ex Claravis pretiosa) (54.7%) #36
PE (C. passerinae 2 ex Claravis pretiosa) (54.7%) #35
PE (C. clayae ex Treron wallia) (54.8%) #6
PE (C. claytoni ex Ducula rufigaster) (54.8%) #7
PE (C. claytoni ex Ducula rufigaster) (54.8%) #8
59/ 0.57
60/ 0.64
53/ 0.56
67/ 0.71
76/0.78
69/ 0.74
86/ 0.88
90/1.0
B
C
A

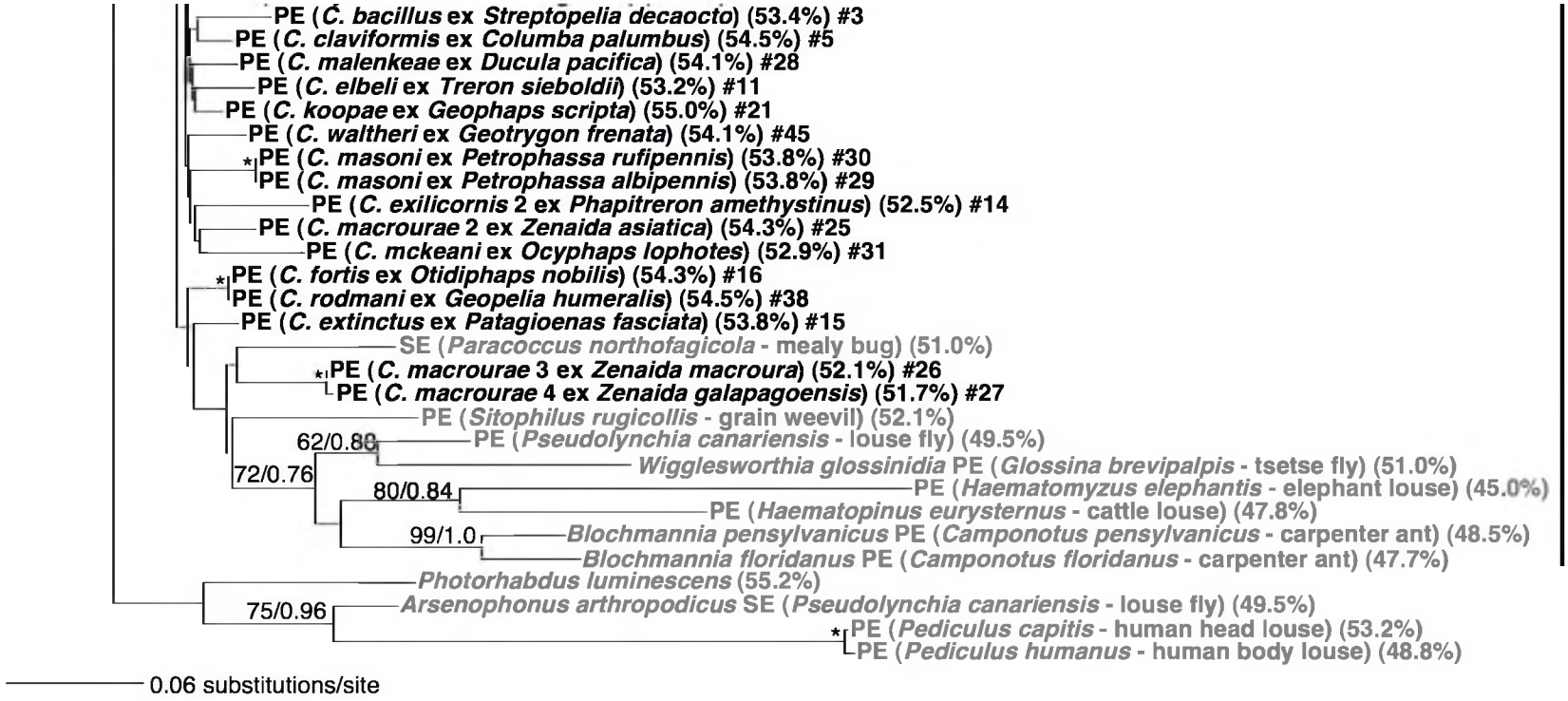


Fig. C1 continued

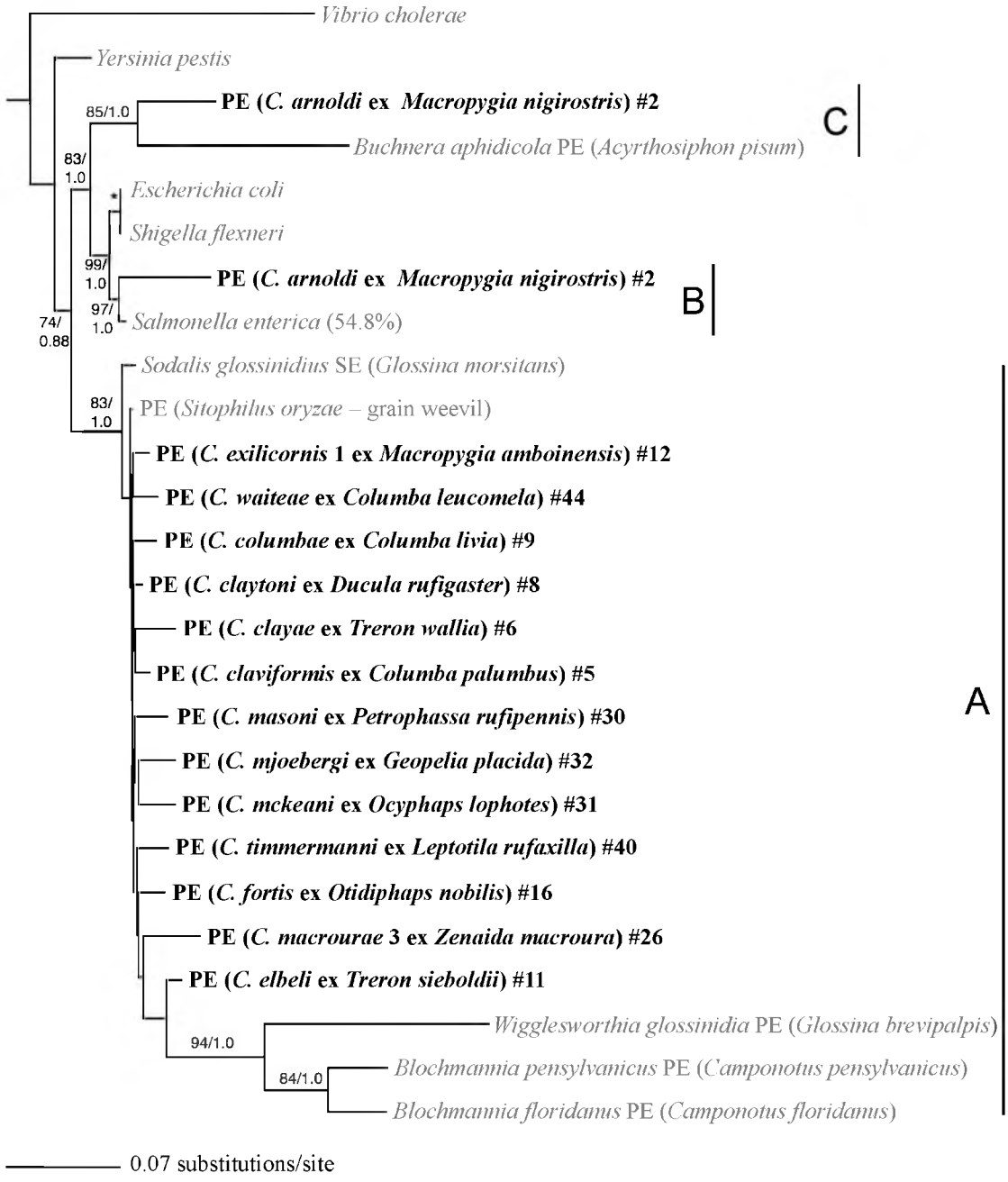


Figure C.2. Phylogeny of *Columbicola* spp. symbionts derived from maximum likelihood and Bayesian analyses of a combined data set consisting of 16S rRNA, *fusA* and *groEL* gene sequences. Conventions as in Fig. C.1.

Results

Identification of Columbicola spp. symbionts

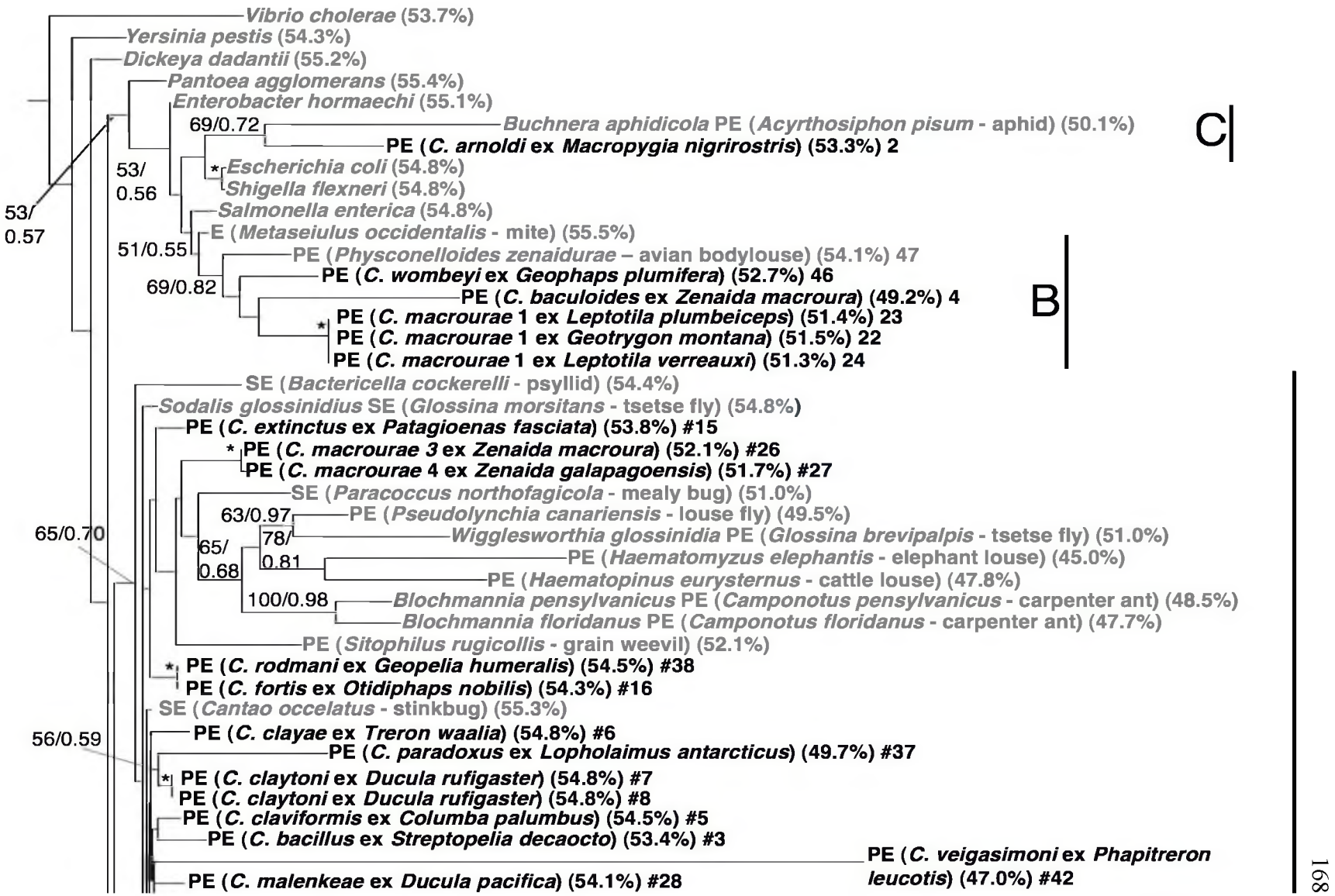
We initially sequenced 48 16S rRNA gene clones from individuals of *C. columbae* and *C. baculoides*. Despite sequencing this many individuals, we obtained only one 16S rRNA gene sequence from each of the two *Columbicola* species. For each of the other *Columbicola* spp. in the study, we sequenced a minimum of four 16S rRNA gene clones. No within species sequence heterogeneity was observed, indicating that each of the *Columbicola* spp. screened in this study is associated with a single bacterial symbiont.

Structural analysis of 16S rRNA sequences of Columbicola spp. symbionts

In our initial 16S rRNA gene phylogeny containing all of the sequences derived from the symbionts of *Columbicola* spp., the sequences from *C. veigasimoni* and *C. paradoxus* exhibited unusually long branches, indicating substantially higher evolutionary rates than the sequences of the other *Columbicola* spp. symbionts of the same clade (Fig. C.3). In addition, the *C. veigasimoni* and *C. paradoxus* symbiont 16S rRNA sequences exhibited unusually low G+C contents relative to the other members of the same clade (Fig. C.3). These patterns suggest the possibility that these highly divergent sequences might represent non-functional copies of the 16S rRNA gene in these symbiont genomes.

Secondary structure analyses of the 16S rRNA sequences using a homology model (Pei et al., 2010) revealed that the *C. veigasimoni* and *C. paradoxus* symbiont 16S rRNA sequences exhibit unusually high ratios of disruptive/conservative nucleotide

Figure C.3. Phylogeny of *Columbicola* spp. symbionts (bold) and related bacteria based on maximum likelihood and Bayesian analyses of a 1.46-kbp fragment of the 16S rRNA gene sequence. Insect symbionts are designed by the prefix "PE" (primary endosymbiont), "SE" (secondary endosymbiont) or E (if unknown), followed by host name and common name (for those not derived from *Columbicola* spp.) The numbers adjacent to nodes indicate maximum likelihood bootstrap values (above the line) and Bayesian posterior probabilities (below the line), for nodes with bootstrap support >50% and Bayesian posterior probabilities >0.5. Asterisks indicate nodes with 100 % bootstrap support and Bayesian posterior probability = 1. Numbers in parentheses represent the G+C content of the 16S rRNA sequences. Final numbers correspond to the list provided in Table C.1.


Vibrio cholerae (53.7%)
Yersinia pestis (54.3%)
Dickeya dadantii (55.2%)
Pantoea agglomerans (55.4%)
Enterobacter hormaechi (55.1%)
69/0.72
Buchnera aphidicola PE (Acyrthosiphon pisum - aphid) (50.1%)
PE (C. arnoldi ex Macropygia nigrirostris) (53.3%) 2
53/
0.56
*
Escherichia coli (54.8%)
Shigella flexneri (54.8%)
53/
0.57
Salmonella enterica (54.8%)
E (Metaseiulus occidentalis - mite) (55.5%)
51/0.55
PE (Physconelloides zenaidurae – avian bodylouse) (54.1%) 47
PE (C. wombeyi ex Geophaps plumifera) (52.7%) 46
69/0.82
PE (C. baculoides ex Zenaida macroura) (49.2%) 4
*
PE (C. macrourae 1 ex Leptotila plumbeiceps) (51.4%) 23
PE (C. macrourae 1 ex Geotrygon montana) (51.5%) 22
PE (C. macrourae 1 ex Leptotila verreauxi) (51.3%) 24
C
B
SE (Bactericella cockerelli - psyllid) (54.4%)
Sodalis glossinidius SE (Glossina morsitans - tsetse fly) (54.8%)
PE (C. extinctus ex Patagioenas fasciata) (53.8%) #15
*
PE (C. macrourae 3 ex Zenaida macroura) (52.1%) #26
PE (C. macrourae 4 ex Zenaida galapagoensis) (51.7%) #27
SE (Paracoccus northofagicola - mealy bug) (51.0%)
63/0.97
PE (Pseudolynchia canariensis - louse fly) (49.5%)
65/0.70
78/
0.81
Wigglesworthia glossinidia PE (Glossina brevipalpis - tsetse fly) (51.0%)
65/
0.68
PE (Haematomyzus elephantis - elephant louse) (45.0%)
PE (Haematopinus eurysternus - cattle louse) (47.8%)
100/0.98
Blochmannia pensylvanicus PE (Camponotus pensylvanicus - carpenter ant) (48.5%)
Blochmannia floridanus PE (Camponotus floridanus - carpenter ant) (47.7%)
PE (Sitophilus rugicollis - grain weevil) (52.1%)
*
PE (C. rodmani ex Geopelia humeralis) (54.5%) #38
PE (C. fortis ex Otidiphaps nobilis) (54.3%) #16
SE (Cantao occelatus - stinkbug) (55.3%)
56/0.59
PE (C. clayae ex Treron waalia) (54.8%) #6
PE (C. paradoxus ex Lopholaimus antarcticus) (49.7%) #37
*
PE (C. claytoni ex Ducula rufigaster) (54.8%) #7
PE (C. claytoni ex Ducula rufigaster) (54.8%) #8
PE (C. claviformis ex Columba palumbus) (54.5%) #5
PE (C. bacillus ex Streptopelia decaocto) (53.4%) #3
PE (C. veigasimoni ex Phapitreron leucotis) (47.0%) #42
PE (C. malenkeae ex Ducula pacifica) (54.1%) #28

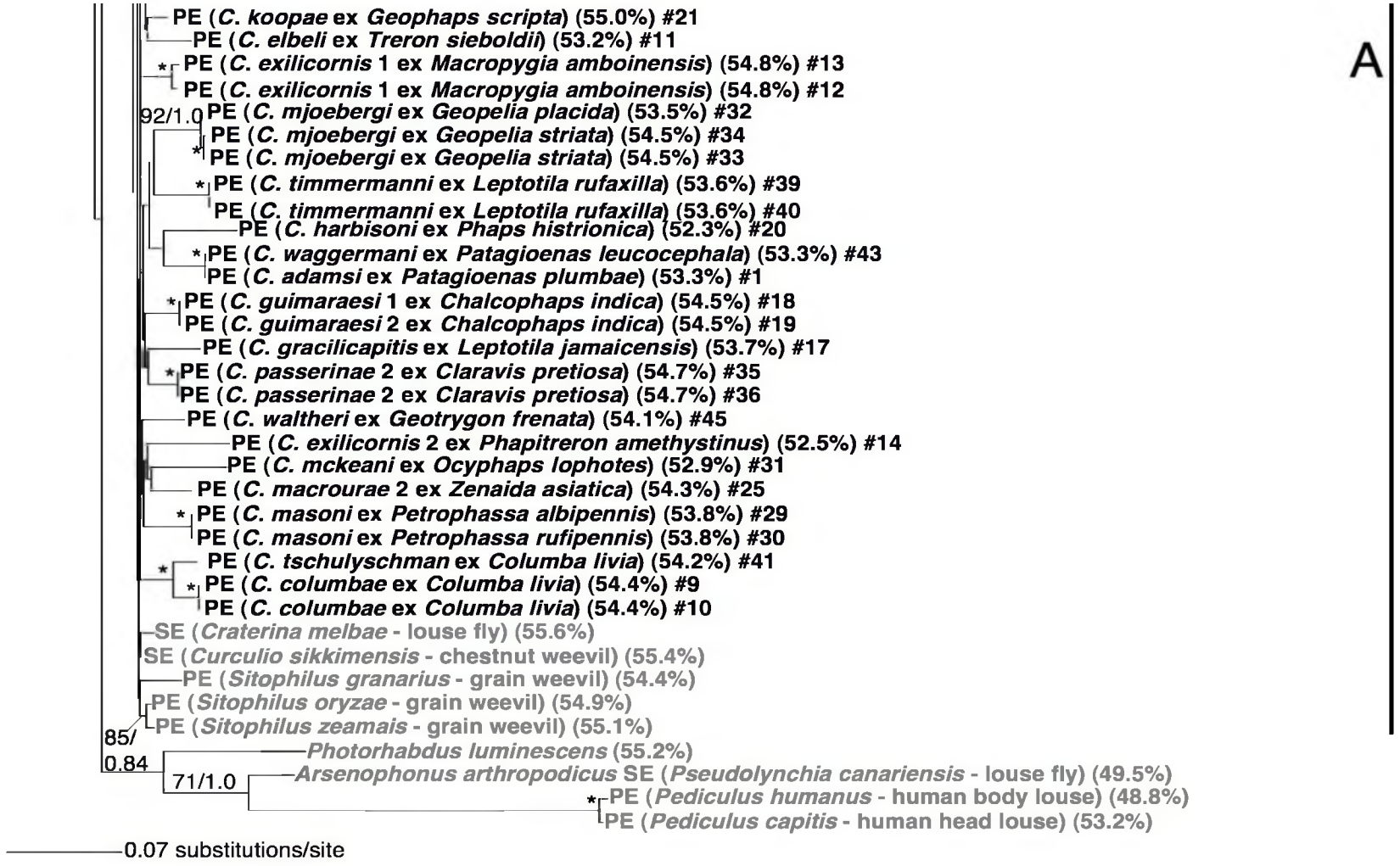
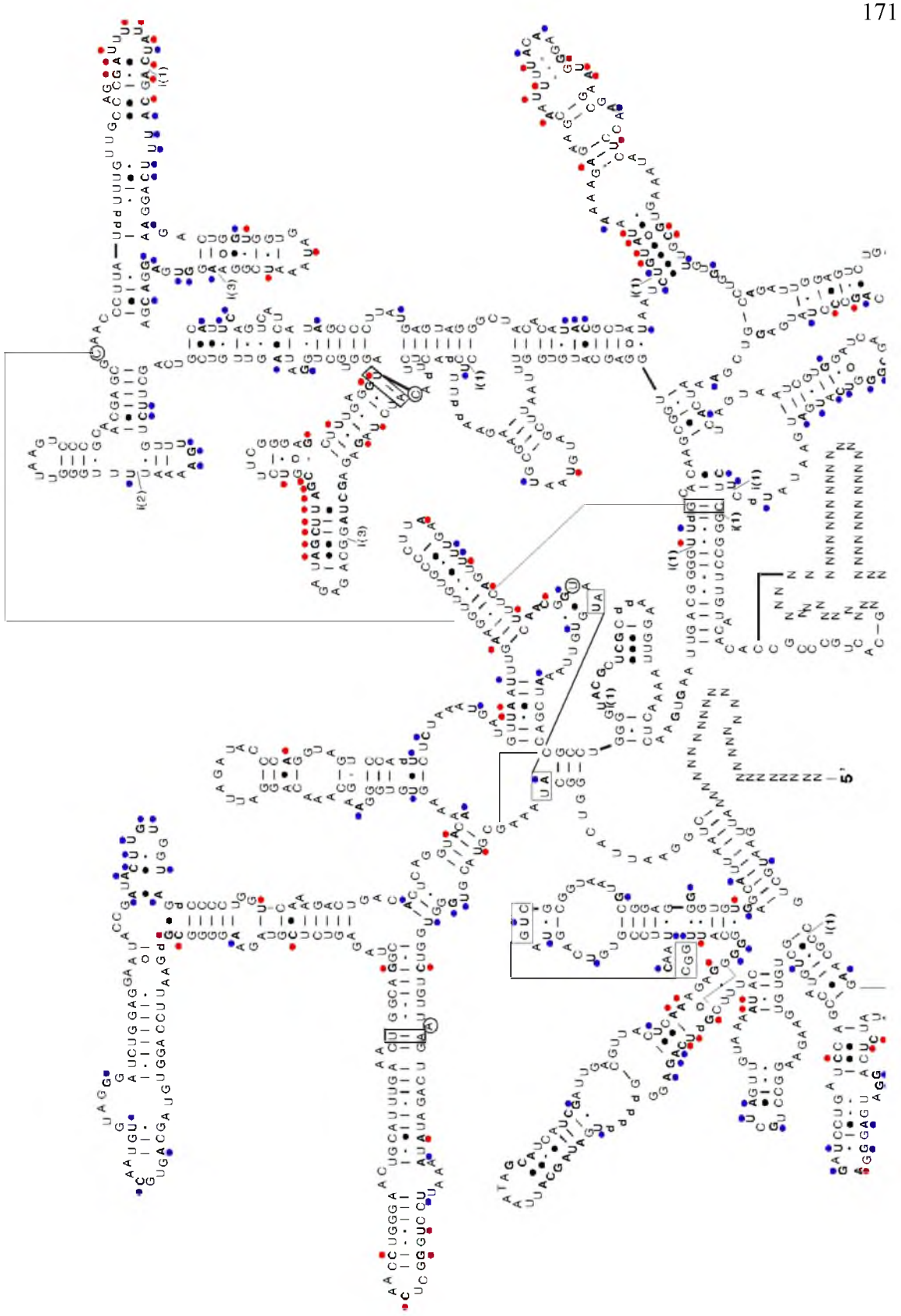


Fig. C3 continued

Figure C.4. Homology model depicting the *C. veigasimoni* symbiont 16S rRNA sequence mapped onto the predicted *Y. pestis* 16S rRNA structure. Homology was deduced from an alignment generated in Muscle, and adjusted manually to account for indels. Substitutions in the symbiont 16S rRNA are highlighted in bold. Substitutions with a higher-than-average rate of variability ($v > 1$) are highlighted with red spots, whereas those with a lower-than-average rate of variability ($v < 1$) are highlighted with blue spots. The counts of different substitution types are displayed in parentheses in the key.



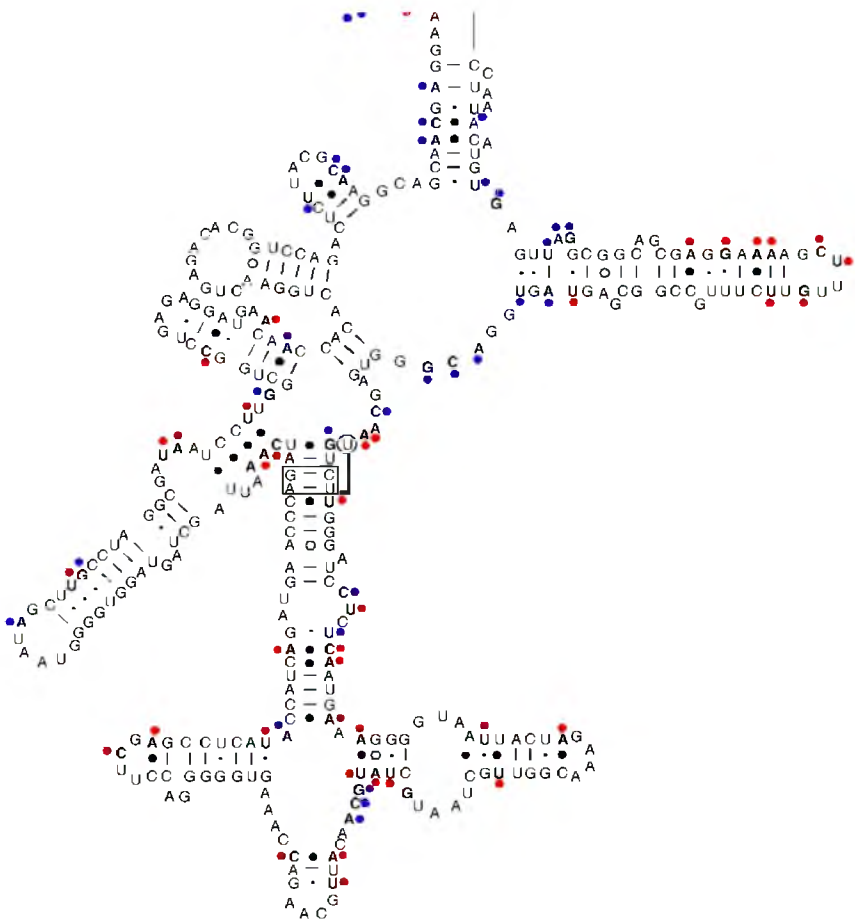
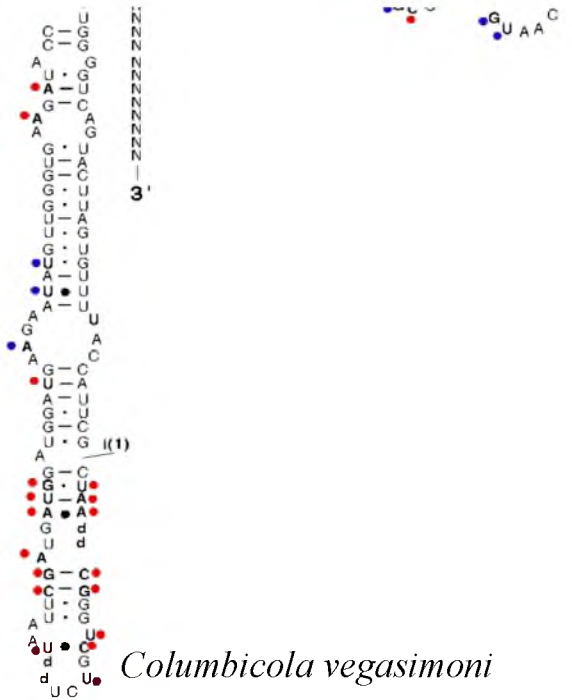


Fig. C4 continued



Columbicola vegasimoni

i Insertion with number of bases in parentheses

d Deletion

● Disrupting non-canonical pairing (n=85)

— Canonical pairing (n=48)

○ Conservative non-canonical G:A pairing (n=6)

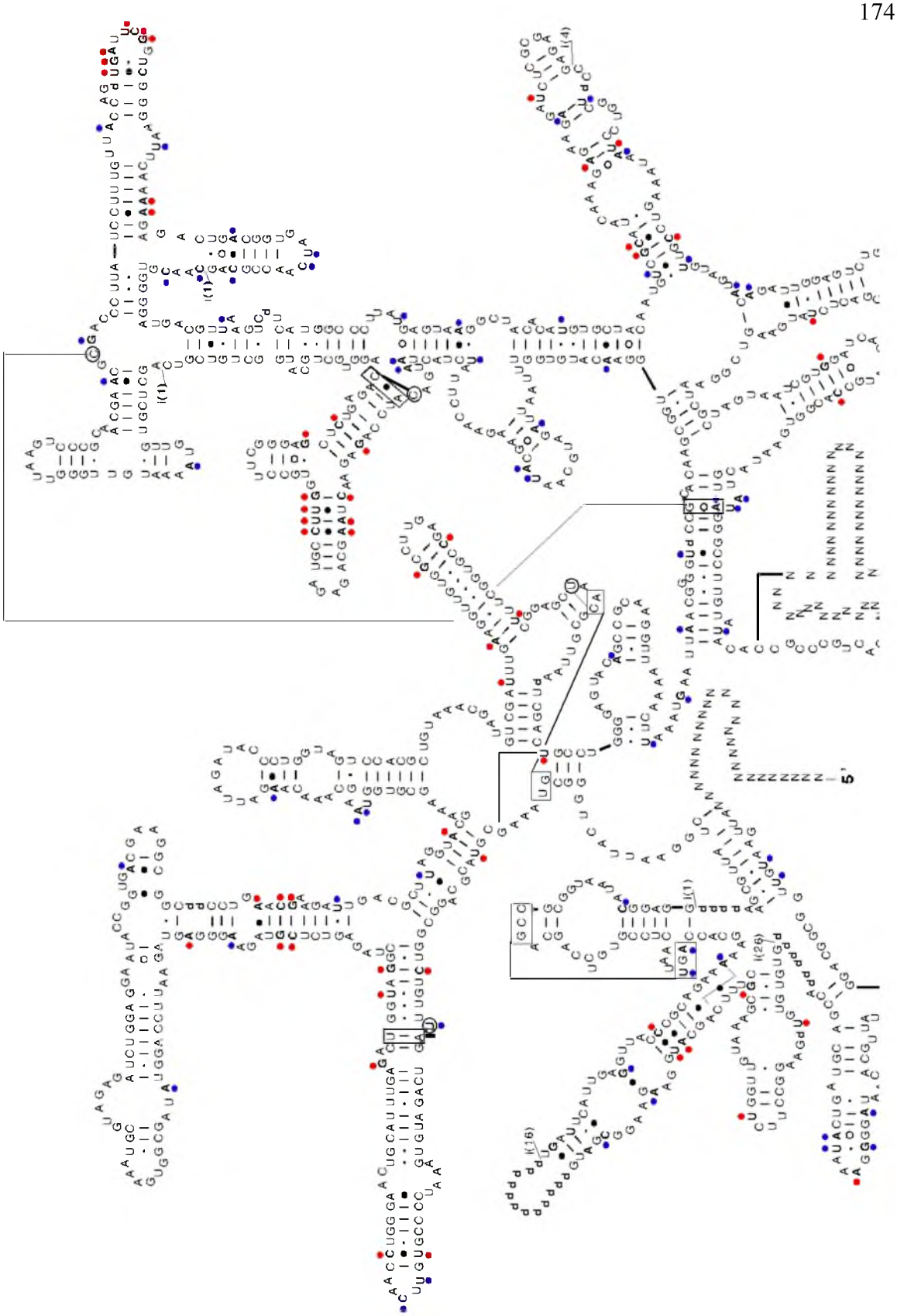
• Conservative non-canonical G:U pairing (n=41)

● Site with variability below average for all sites (v<1) (n=148)

● Site with variability above average for all sites (v>1) (n=137)

Probable tertiary interactions are connected with solid lines

Figure C.5. Homology model depicting the *C. paradoxus* symbiont 16S rRNA sequence mapped onto the predicted *Y. pestis* 16S rRNA structure. Conventions are as in Figure C.4.



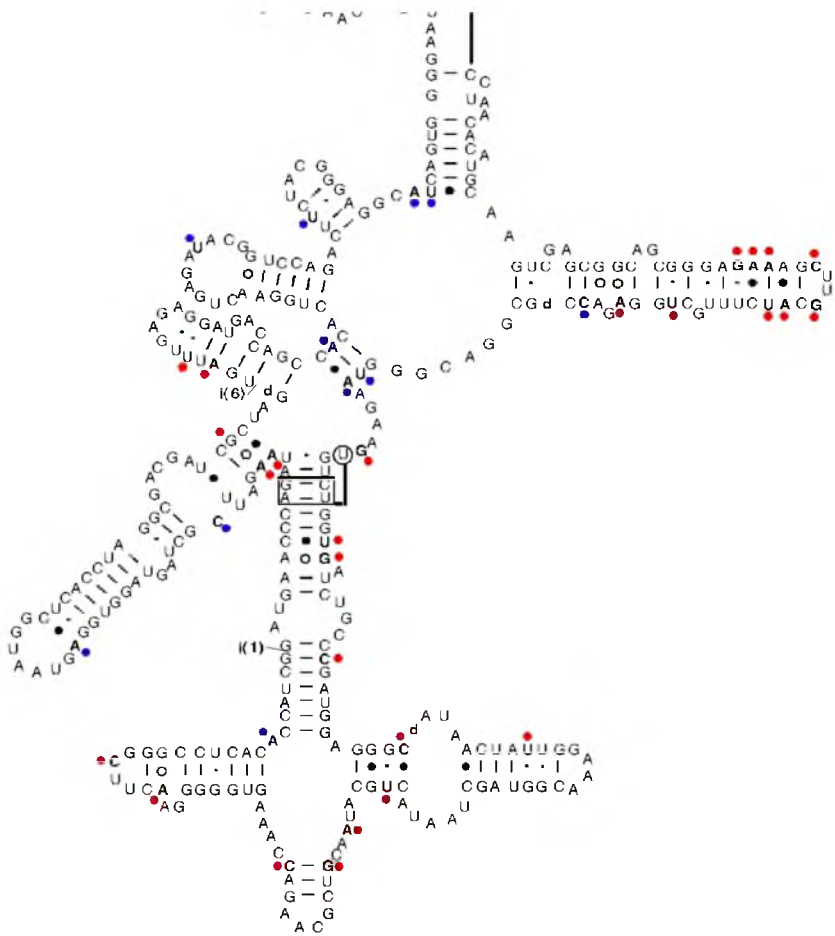
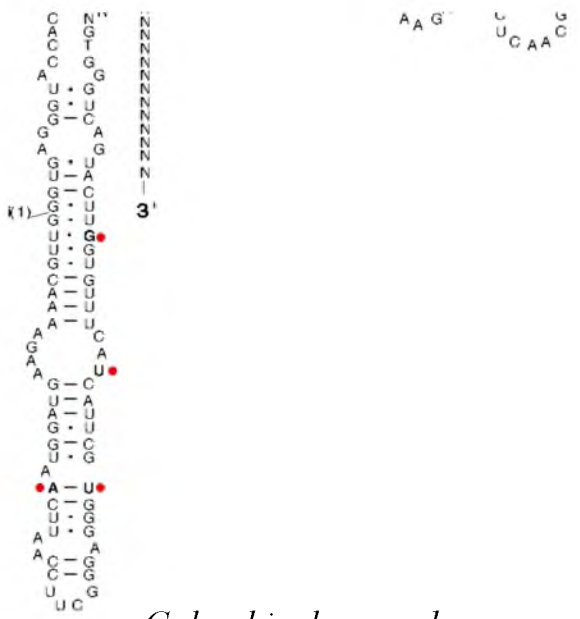


Fig. C5 continued



Columbicola paradoxus

i Insertion with number of bases in parentheses

d Deletion

● Disrupting non-canonical pairing (n=30)

— Canonical pairing (n=40)

○ Conservative non-canonical G:A pairing (n=8)

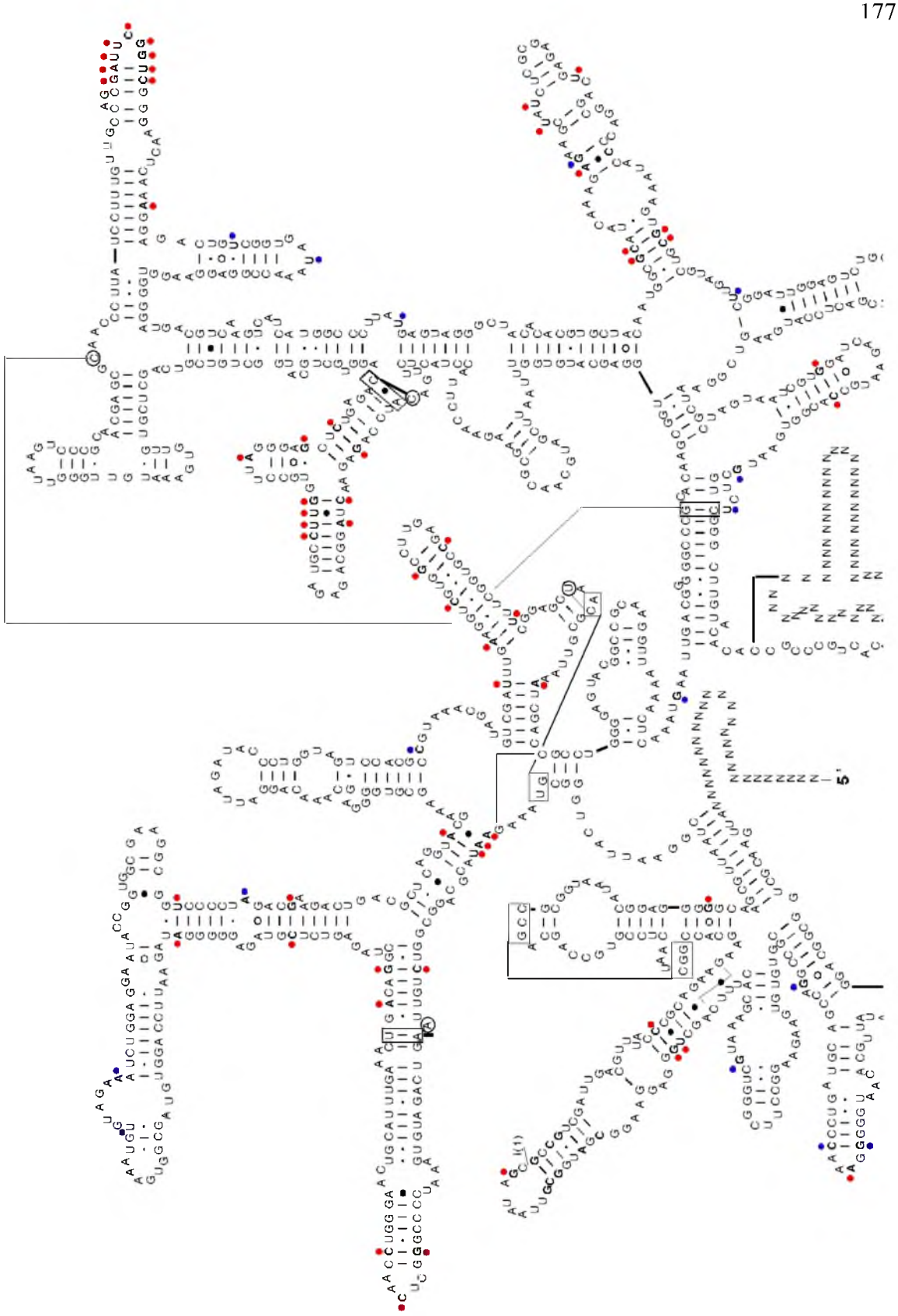
• Conservative non-canonical G:U pairing (n=20)

● Site with variability below average for all sites (v<1) (n=70)

● Site with variability above average for all sites (v>1) (n=84)

Probable tertiary interactions are connected with solid lines

Figure C.6. Homology model depicting the *C. columbae* symbiont 16S rRNA sequence mapped onto the predicted *Y. pestis* 16S rRNA structure. Conventions are as in Figure C.4.



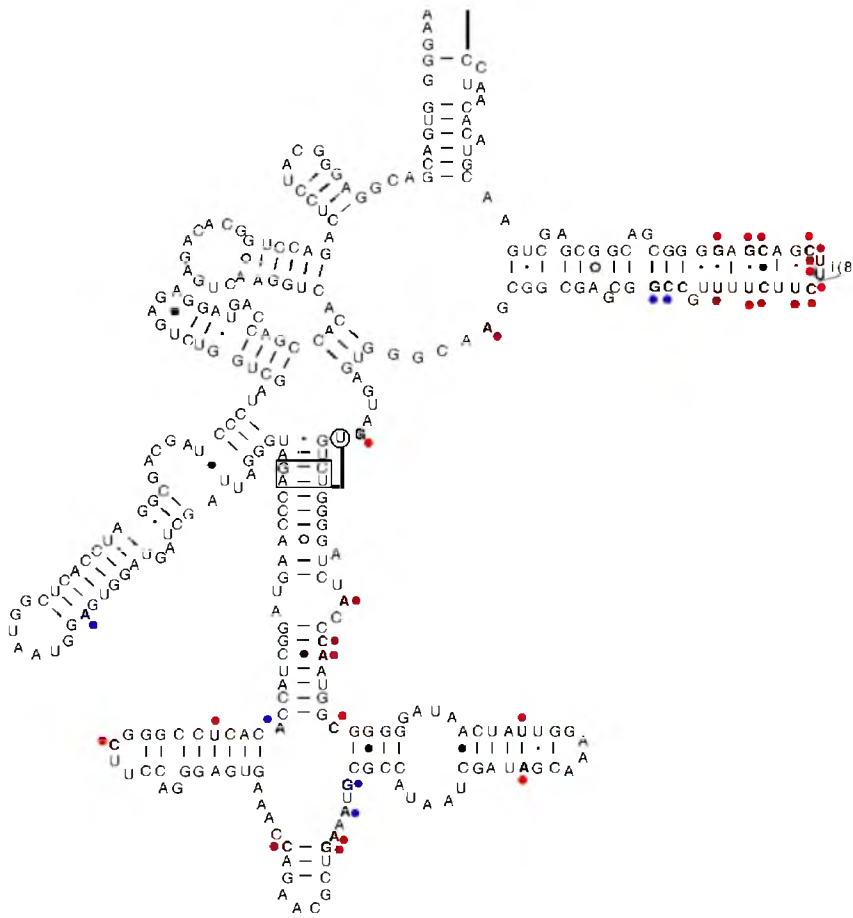
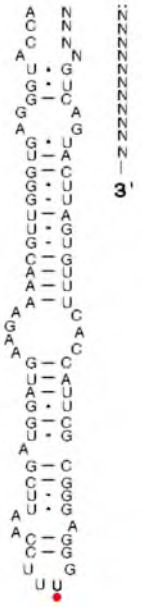


Fig. C6 continued



Columbicola columbae

i Insertion with number of bases in parentheses

d Deletion

● Disrupting non-canonical pairing (n=9)

— Canonical pairing (n=50)

○ Conservative non-canonical G:A pairing (n=3)

• Conservative non-canonical G:U pairing (n=7)

● Site with variability below average for all sites (v<1) (n=22)

● Site with variability above average for all sites (v>1) (n=86)

Probable tertiary interactions are connected with solid lines

substitutions. In the *C. veigasimoni* 16S rRNA sequence, 85 out of a total of 180 substitutions (47.2%) were predicted to encode disruptive changes (causing putative stem-loop transitions; Fig. C.4). In the *C. paradoxus* 16S rRNA sequence, similarly disruptive substitutions comprised 30 out of a total of 98 substitutions (30.6%; Fig. C.5). In the analysis of the *C. columbae* 16S rRNA sequence residing on a relatively short branch, by contrast, only 9 out of 69 substitutions (13%) were classified as disruptive (Fig. C.6). The ratios of disruptive/conservative substitutions in both the *C. veigasimoni* and *C. paradoxus* symbiont 16S rRNA sequences were significantly higher than in the *C. columbae* symbiont 16S rRNA sequence (Fisher's exact test; $P < 0.001$). By superimposing these data onto the variability map of Wuyts et al., (2001), the *C. veigasimoni* and *C. paradoxus* symbiont 16S rRNA sequences were shown to contain significantly more substitutions in sites that typically display low variability (Fisher's exact test; $P < 0.0001$). All these data strongly suggest that the *C. veigasimoni* and *C. paradoxus* symbiont 16S rRNA sequences are not evolving in accordance with the functional constraints affecting other 16S rRNA sequences, including that of the *C. columbae* symbiont. While the possibility cannot be ruled out that these highly disrupted 16S rRNA sequences are functional, an alternative possibility is that the *C. veigasimoni* and *C. paradoxus* symbionts maintain additional (functional) paralogous copies of 16S rRNA that were not amplified by the universal primers used in this study. On the basis of these results and considerations, we elected to exclude the *C. veigasimoni* and *C. paradoxus* symbiont 16S rRNA gene sequences from the subsequent molecular phylogenetic analyses.

Phylogenetic analysis of 16S rRNA gene sequences of Columbicola spp. symbionts

On the basis of 16S rRNA gene sequences, the symbionts of *Columbicola* spp. formed three distinct clades in the *Gammaproteobacteria* (Fig. C.1). Clade A contains the largest number of *Columbicola* spp. symbionts and is represented by the symbiont of *C. columbae*, and also the tsetse fly symbiont *Sodalis glossinidius* and symbionts of grain weevils of the genus *Sitophilus*. Clade B is represented by several *Columbicola* spp. symbionts, including the symbiont of *C. baculoides*, and symbionts identified from the avian body louse *Physconelloides zenaidurae* and the mite *Metaseiulus occidentalis* (Hoy and Jeyaprakash, 2005). The sole representative of clade C, the symbiont of *C. arnoldi*, is distantly allied to the aphid primary endosymbiont *Buchnera aphidicola*.

Phylogenetic analysis of multiple gene sequences of Columbicola spp. symbionts

On the basis of combined sequence data of 16S rRNA, *fusA* and *groEL* genes, the symbionts of *Columbicola* spp. were also divided into three distinct clades A, B and C in the *Gammaproteobacteria* (Fig. C.2). Here the number of analyzed taxa was smaller because all three genes were not always successfully amplified by PCR from the louse samples, but the phylogenetic relationships were entirely concordant with the analysis of the 16S rRNA sequence data alone (Fig. C.1).

Star-like phylogeny in the clade A of Columbicola spp. symbionts

In the 16S rRNA gene phylogeny (Fig. C.1), the symbiont sequences from different individuals of the same species/cryptic species/haplogroups usually formed

monophyletic groups with high statistical support Meanwhile, deeper relationships between the symbionts of *Columbicola* spp. were not well resolved regardless of the reconstruction method employed. In particular, the internodes connecting the representatives of clade A were extremely short with little or no statistical support, although substantial sequence divergence was observed among the representatives of clade A, as evidenced by the relatively long branches leading to terminal nodes. In clade A, consequently, the phylogeny exhibited a comb- or star-like appearance, except for the following statistically-supported terminal clusters of recent origin: (i) *C. fortis* #16 and *C. rodmani* #38; (ii) *C. adamsi* #1 and *C. waggermani* #43; and (iii) *C. columbae* #9, 10 and *C. tschulyshman* #41 (Fig. C.1). In the phylogeny based on the combined 16S rRNA, *fusA* and *groEL* gene sequences (Fig. C.2), no clusters with significant statistical support were identified in clade A, which contained consistently long terminal branches and corroborated the star-like phylogenetic relationship in the clade A symbionts of *Columbicola* spp.

Nucleotide composition

In the 16S rRNA gene phylogeny (Fig. C.1), free-living gammaproteobacterial representatives generally exhibited G+C contents around 55%. The values of the clade A symbionts ranged mostly from 53% to 55%, although several symbiont lineages exhibited lower values, such as those from *C. macrourae* #3, 4 (51.7%-52.1%), *C. exilicornis* #20 (52.5%) and *C. harbisoni* #20 (52.3%). The G+C contents of the clade B symbionts were split into two groups: low values for the symbionts of *C. baculoides* #4 (49.2%) and *C. macrourae* #22-24 (51.3%-51.5%), and high values for the symbionts of *C. wombeyi* #46

(52.7%) and *P. zenaidurae* #47 (54.1%). The G+C content of the sole clade C symbiont of *C. arnoldi* #2 was 53.3%.

Molecular evolutionary rate estimations

Relative rate tests revealed that the substitution rates in 16S rRNA genes of the clade A (*C. columbae*), clade B (*C. baculoides*) and clade C (*C. arnoldi*) symbionts were significantly higher than those of their free-living relatives. Clade B and clade C symbionts exhibited 1.42 times and 1.35 times higher rates, respectively, which were supported by highly significant ($\sim 10^{-5}$) *P*-values, while the clade A symbiont showed a 1.17 times higher rate, which was supported by a moderately sgnificant ($\sim 10^{-2}$) *P*-value (Table C.3).

In vivo localization of C. baculoides symbiont

In males of *C. baculoides*, fluorescent *in situ* hybridization detected the symbiont cells within bacteriocytes that clustered on both sides of the abdominal cavity (Fig. C.7A, B and C). In young females of *C. baculoides*, the symbiont cells exhibited the same localization as in males (not shown). In mature females, by contrast, the symbiont cells were found in ovarial tissues, localized in a pair of specialized transmission organs called ovarial ampullae (Fig. C.7D and E), and vertically transmitted from the ovarial ampullae to the posterior pole of developing oocytes (Fig. C.7F). The localization, migration and transmission patterns of the clade B symbiont of *C. baculoides* are almost identical to those of the clade A symbiont of *C. columbae* (Fukatsu et al., 2007).

Table C.3. Relative-rate tests comparing molecular evolutionary rates of 16S rRNA gene sequences between different lineages of the symbionts of *Columbicola* spp. and free-living relatives.

Lineage 1	**Lineage 2**	**Outgroup**	**K1**[1]	**K2**[2]	**K1-K2**	**K1/K2**	***P*-value**[3]
Symbiont of *C. columbae* (clade A) [JQ063426]	*Dickeya dadantii* [AY360397]	*Vibrio cholerae* [CP001486.1]	0.124	0.106	0.018	1.170	0.012
Symbiont of *C. baculoides* (clade B) [JQ063412]	*Salmonella enterica* [AP011957]	*Vibrio cholerae* [CP001486.1]	0.143	0.101	0.042	1.416	2.19×10^{-5}
Symbiont of *C. arnoldi* (clade C) [JQ963434]	*Escherichia coli* [U00096]	*Vibrio cholerae* [CP001486.1]	0.142	0.105	0.037	1.352	9.18×10^{-6}
Symbiont of *C. columbae* (clade A) [JQ063426]	Symbiont of *C. baculoides* (clade B) [JQ063412]	*Vibrio cholerae* [CP001486.1]	0.124	0.143	-0.019	0.867	0.105
Symbiont of *C. baculoides* (clade B) [JQ063412]	Symbiont of *C. arnoldi* (clade C) [JQ963434]	*Vibrio cholerae* [CP001486.1]	0.143	0.142	0.001	1.007	0.981
Symbiont of *C. arnoldi* (clade C) [JQ963434]	Symbiont of *C. columbae* (clade A) [JQ063426]	*Vibrio cholerae* [CP001486.1]	0.142	0.124	0.018	1.145	0.065
Symbiont of *C. columbae* (clade A) [JQ063426]	*Sodalis glossinidius* (clade A) [AF548135]	*Vibrio cholerae* [CP001486.1]	0.124	0.106	0.018	0.169	0.001
Sodalis glossinidius (clade A) [AF548135]	*Dickeya dadantii* [AY360397]	*Vibrio cholerae* [CP001486.1]	0.106	0.106	0	0	0.998

[1]Estimated mean distance between lineage 1 and the last common ancestor of lineages 1 and 2.
[2]Estimated mean distance between lineage 2 and the last common ancestor of lineages 1 and 2.
[3]*P*-value was generated using the program RRTree (Robinson-Rechavi and Huchon 2000).

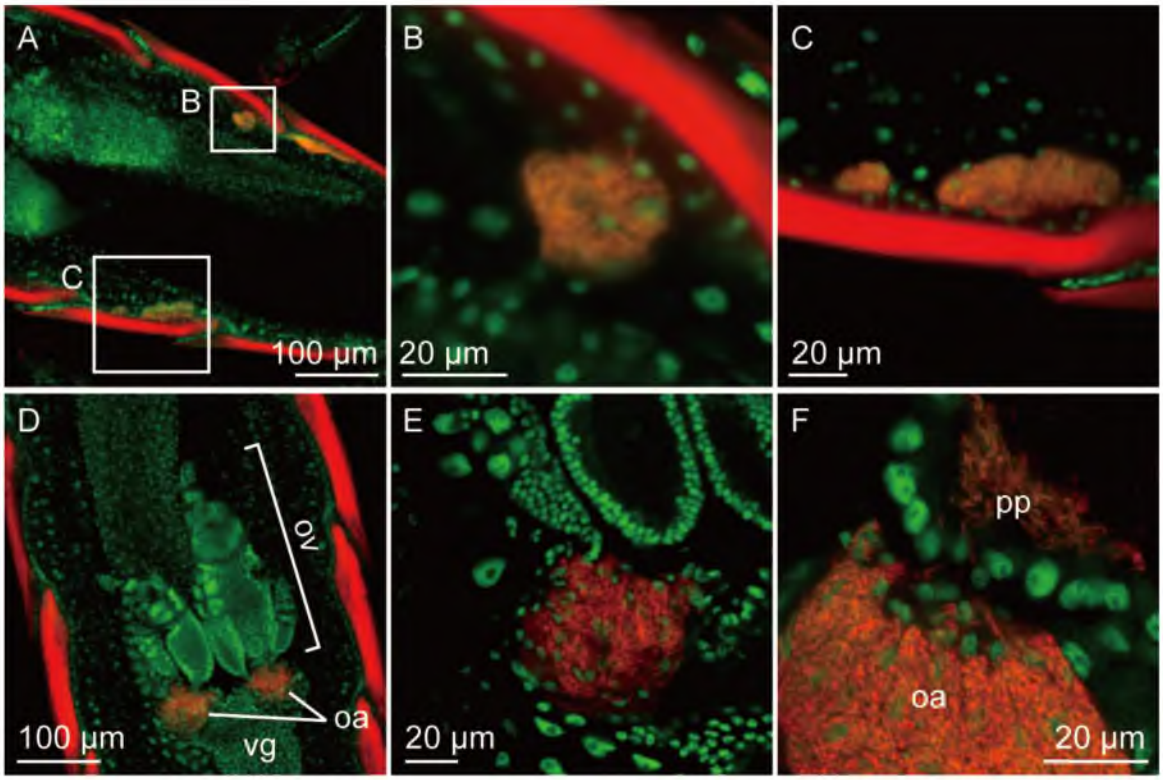


Figure C.7. Fluorescent *in situ* hybridization of the symbiont in *C. baculoides*. (A) Abdominal image of an adult male. Signals of the symbiont cells are detected in bacteriocyte clusters located on both sides of the abdominal body cavity. White squares indicate the areas of panels B and C. (B, C) Enlarged images of the bacteriocyte clusters in panel A. Signals of the symbiont cells are localized in the cytoplasm of the bacteriocytes. (D) Abdominal image of an adult female. Signals of the symbiont cells are localized in ovarial ampullae located at the base of the ovaries. (E) An enlarged image of an ovarial ampulla. (F) A snapshot of symbiont transmission from an ovarial ampulla to a developing oocyte. Red and green signals indicate symbiont 16S rRNA and host nuclear DNA, respectively. Abbreviations: oa: ovarial ampulla, ov: ovariol, vg: vagina, pp: posterior pole of oocyte.

Lack of host-symbiont phylogenetic congruence

Using the S-H test (Shimodaira and Hasegawa 1999), the phylogenetic relationship of *Columbicola* spp. (Johnson et al., 2007) was statistically rejected by the symbiont 16S rRNA gene dataset alone (difference in ln L = 575.24, $P < 0.001$) and by the combined dataset of 16S rRNA, *fusA* and *groEL* genes (difference in ln L = 46.43, $P < 0.001$).

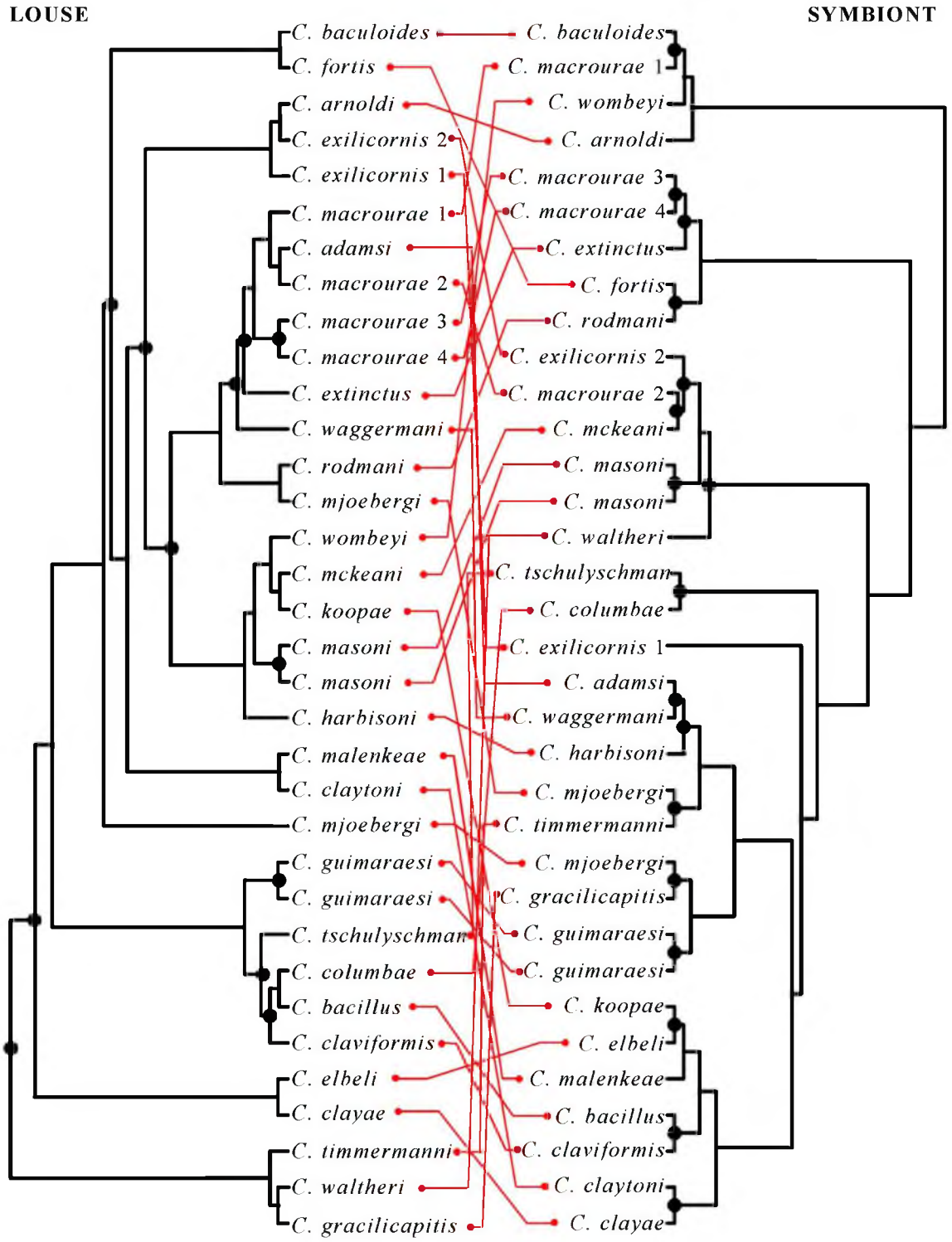
Co-phylogenetic analysis (Page 1995) of the 16S rRNA gene dataset reconstructed 17 potential co-speciation events between the host and symbiont lineages (Fig. C.8A). However, this number of co-speciation events was not significantly higher than expected by chance ($P > 0.05$). Using the combined dataset of 16S rRNA, *fusA* and *groEL* genes, co-phylogenetic analysis reconstructed only 6 potential co-speciation events (Fig. C.8B), again no more than that expected by chance ($P > 0.05$).

Discussion

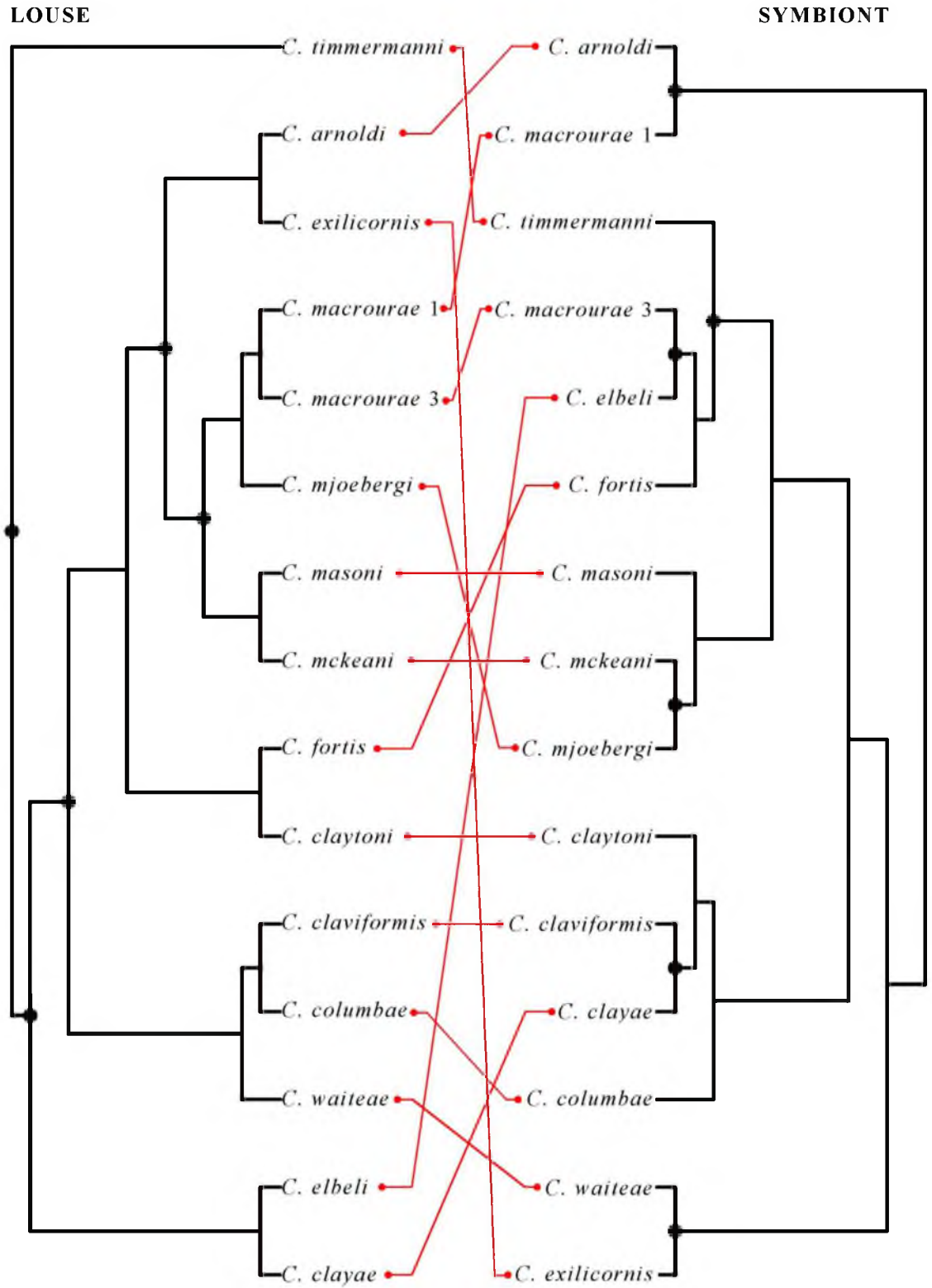
We undertook an extensive molecular survey of bacterial symbionts associated with feather-feeding lice of the genus *Columbicola*, which are found on most species of pigeons and doves across the globe. The molecular phylogenetic analyses placed the symbionts of *Columbicola* spp. into three clades, designated A, B and C, in the *Gammaproteobacteria*. Each of the representative louse samples screened in this study was associated with a single bacterial symbiont representing one of the three clades. The same pattern was consistently observed in multiple individuals of the same louse species. From these results, we conclude that each of the columbiform feather lice of the genus *Columbicola* is associated with a specific primary bacterial symbiont, while the symbionts associated with related species of *Columbicola* may be phylogenetically

Figure C.8. Comparison of the phylogenies of representative species of *Columbicola* spp. and their symbiotic bacteria. *Columbicola* trees are from maximum likelihood analysis of sequences of the mitochondrial cytochrome oxidase I gene, mitochondrial 12S rRNA gene, and nuclear elongation factor 1alpha gene from Johnson *et al.* (2007). Symbiont trees are from Fig. C.1 (A) and Fig. C.2 (B) in the current paper. Connecting lines illustrate host-symbiont associations. Bulleted nodes are co-speciation events inferred from reconciliation analysis (Page 1995).

A


LOUSE
SYMBIONT
C. baculoides
C. fortis
C. arnoldi
C. exilicornis 2
C. exilicornis 1
C. macrourae 1
C. adamsi
C. macrourae 2
C. macrourae 3
C. macrourae 4
C. extinctus
C. waggermani
C. rodmani
C. mjoebergi
C. wombeyi
C. mckeani
C. koopae
C. masoni
C. masoni
C. harbisoni
C. malenkeae
C. claytoni
C. mjoebergi
C. guimaraesi
C. guimaraesi
C. tschulyschman
C. columbae
C. bacillus
C. claviformis
C. elbeli
C. clayae
C. timmermanni
C. waltheri
C. gracilicapitis
C. baculoides
C. macrourae 1
C. wombeyi
C. arnoldi
C. macrourae 3
C. macrourae 4
C. extinctus
C. fortis
C. rodmani
C. exilicornis 2
C. macrourae 2
C. mckeani
C. masoni
C. masoni
C. waltheri
C. tschulyschman
C. columbae
C. exilicornis 1
C. adamsi
C. waggermani
C. harbisoni
C. mjoebergi
C. timmermanni
C. mjoebergi
C. gracilicapitis
C. guimaraesi
C. guimaraesi
C. koopae
C. elbeli
C. malenkeae
C. bacillus
C. claviformis
C. claytoni
C. clayae

B


LOUSE
SYMBIONT
C. timmermanni
C. arnoldi
C. exilicornis
C. macrourae 1
C. macrourae 3
C. mjoebergi
C. masoni
C. mckeani
C. fortis
C. claytoni
C. claviformis
C. columbae
C. waiteae
C. elbeli
C. clayae
C. arnoldi
C. macrourae 1
C. timmermanni
C. macrourae 3
C. elbeli
C. fortis
C. masoni
C. mckeani
C. mjoebergi
C. claytoni
C. claviformis
C. clayae
C. columbae
C. waiteae
C. exilicornis

distinct, implying their polyphyletic evolutionary origins.

In many obligate host-symbiont associations, such as those observed between aphids and *Buchnera aphidicola* and tsetse flies and *Wigglesworthia glossinidia*, a specific host taxon is associated with a single evolutionary origin of the symbiont in the common ancestor of the host clade, followed by stable host-symbiont association and repeated bouts of co-speciation, yielding congruent phylogenies (Baumann, 2005; Moranet al., 2008). In this study we demonstrated that evolutionary patterns of host symbiont association are quite different in the *Columbicola* feather lice: the primary symbionts are of polyphyletic evolutionary origins and do not exhibit host-symbiont co-speciation. Similar polyphyletic primary symbionts have been reported from weevils of the family Dryophthoridae (Lefevre et al., 2004; Conord et al., 2008), and sucking lice of the suborder Anoplura (Hypsa and Krizek, 2007). However, the case of *Columbicola* spp. is remarkable in that such polyphyletic primary symbionts are found within a single genus of host insects.

Relative rate tests revealed that molecular evolutionary rates are elevated in representatives of all three symbiont clades associated with *Columbicola* spp., but the levels of acceleration are more pronounced in clades B and C relative to clade A (Fig. C.1; Table C.3). These evolutionary patterns suggest that clade B and clade C may be older symbiont lineages that have experienced a longer history of host-symbiont co-evolution and that clade A is of more recent origin. The different symbiont lineages may have been acquired by the different *Columbicola* lineages independently or, alternatively, representatives of clade A may have replaced the older clade B and C lineages. The replacement scenario seems more likely, given that all *Columbicola* species examined in

this study were found to harbor only a single symbiont, and that the clade B symbiont of *C. baculoides* exhibits the same localization, migration and transmission as the clade A symbiont of *C. columbae* (Fig. C.7; Fukatsu et al., 2007).

Previous studies suggested that *Columbicola* feather lice diversified mainly in the Paleogene (Pereira et al., 2007; Johnson and Weckstein, 2011). Hence, replacements of the clade B or C symbionts with representatives of clade A must have taken place during this period. Similar symbiont replacements have been reported from weevils of the family Dryophthoridae, where the ancient symbiont lineage *Nardonella* was replaced by several symbiont lineages that are predicted to be of more recent origin (Lefevre et al., 2004; Conord et al., 2008). Also, in aphids of the tribe Cerataphidini, the ancient symbiont *Buchnera aphidicola* is thought to have been replaced by fungal symbiont lineages (Fukatsu and Ishikawa, 1992; 1996; Fukatsu et al., 1994).

Molecular phylogenetic analyses of feather lice reveal several well-supported clades within the genus *Columbicola* (Johnson et al., 2007). Thus, it is striking that our 16S rRNA gene phylogeny for *Columbicola symbionts* (Fig. C.1) reveals few well-supported clades within the clade A symbionts associated with most of the *Columbicola* species that we examined. Moreover, almost all of the internal branches in clade A are very short, generating a comb- or star-like tree topology (Fig. C.1). Furthermore, co-phylogenetic analyses confirm incongruence between clade A symbionts and their *Columbicola* hosts (Fig. C.8). These results indicate a lack of host-symbiont co-speciation among the clade A symbionts of *Columbicola* spp., providing further support for the notion of horizontal transfers and/or replacements of symbionts within this insect group.

In weevils of the Dryophthoridae, it has been suggested that symbiont replacements might have been driven by major changes in the insect diet (Lefevre et al., 2004). In contrast, all *Columbicola* species are obligate parasites of columbiform birds that live on a diet of feathers and dead skin. Experimental transfers of *Columbicola* spp. between different species of pigeons and doves show that these lice are capable of feeding, surviving and reproducing on feathers of heterospecific hosts (Bush and Clayton, 2006). Hence, symbiont replacements in *Columbicola* spp. are unlikely to be attributable to dietary changes. Previous studies have suggested that biological vectors such as parasitic wasps and mites might facilitate symbiont transfers and replacements across different host species (e.g. Huigens et al., 2004; Jaenike et al., 2007). However, neither parasitoid wasps nor ectoparasitic mites have been reported from *Columbicola* spp. (Johnson and Clayton, 2003). Another possibility is that horizontal symbiont transfers are mediated between different louse species by interspecific mating, as reported for facultative symbionts in the pea aphid (Moran and Dunbar, 2006). Thus far, no direct evidence has been obtained to support this hypothesis, but it is notable that these lice undergo host switching through phoretic dispersion on hippoboscid louse flies (Harbison and Clayton, 2011). In addition, male lice often remain in copula with their female partners for several hours (Johnson and Clayton, 2003), which could provide a possible window for horizontal symbiont transfers.

However, phylogenetic lines of evidence do not necessarily favor the above-mentioned hypothesis of horizontal symbiont transfers between *Columbicola* spp. If different *Columbicola* species had been undergoing occasional symbiont transfers, the resulting symbiont phylogenetic tree would be expected to be of compact shape, with

relatively short terminal branches, as observed for facultative insect symbionts, such as *Wolbachia*, *Rickettsia*, *Hamiltonella*, *Regiella* and *Serratia* (e.g. Werren et al., 1995; Russell et al., 2003; Weinert et al., 2009). Contrary to this expectation, the phylogenies of the symbionts of *Columbicola* spp. are characterized by long terminal branches and very short internodes, giving the trees their comb- or star-like appearance (Fig. C.1 and C.2).

To account for the tree topologies observed in our study we propose an alternative hypothesis that involves repeated symbiont acquisitions from a common "progenitor" bacterial clade that might be ubiquitously present in the environment. This hypothesis is compatible with the following observations and evolutionary patterns: (i) the symbionts of *Columbicola* spp., in particular those of the clade A, are closely related to symbionts of phylogenetically distant insect hosts that encompass diverse geographical and ecological habitats, such as tsetse flies, louse flies, grain weevils, chestnut weevils, longicorn beetles and stinkbugs (Dale and Maudlin, 1999; Heddi and Nardon, 2005; Novakova and Hypsa, 2007; Toju et al., 2010; Grünbald et al., 2010; Kaiwa et al., 2010; 2011; Toju and Fukatsu, 2011); (ii) the symbiont lineages of *Columbicola* spp., tsetse flies and grain weevils, which show ~100% infection frequencies (Dale and Maudlin, 1999; Heddi and Nardon, 2005; Fukatsu et al., 2007), suggestive of relative stability and continuity of the associations, are generally characterized by long terminal branches in the phylogeny (see Fig. C.1); and (iii) by contrast, the symbiont lineages of chestnut weevils and stinkbugs, which show low infection frequencies (Kaiwa et al., 2010; 2011; Toju and Fukatsu, 2011), which may be indicative of instability and/or temporality of the associations, are characterized by very short terminal branches in the phylogeny (see Fig. C.1). We assume that (1) the progenitor bacterial lineage is free-living in the environment with a low rate

of molecular evolution, and (2) establishment of the vertically-transmitted endosymbiotic lifestyle results in accelerated molecular evolution of the symbiont genes (Moran, 1996; Wernegreen, 2002). Given these assumptions, the relatively old *Columbicola* symbiont lineages must have experienced accelerated molecular evolution for longer periods and they therefore tend to exhibit longer terminal branches, whereas the relatively young and/or temporal symbiont lineages have experienced relatively limited evolutionary acceleration and thus exhibit very short terminal branches (Fig. C.9). In order to substantiate this hypothesis, such progenitor lineages will need to be identified in the environment.

Genome sequences of obligate bacterial symbionts are known to degenerate over time as a consequence of their small effective population size (Moran, 1996; Wernegreen, 2002; Moran et al., 2008). Such degenerative evolution may result in tiny symbiont genomes that facilitate replacements via a "revolving door" scenario (Moya et al., 2009). It should be noted, however, that the extremely reduced genomes (< 0.2 Mb) of *Carsonella* and *Tremblaya,* which are endosymbionts of psyllids and mealybugs, respectively, are highly conserved among all host species examined, with no evidence of symbiont replacement (Thao et al., 2000; Thao et al., 2002; Nakabachi et al., 2007; Gruwell et al., 2010; McCutcheon and von Dohlen, 2011). Hence, reductive genome evolution, by itself, cannot be the principal driving force for symbiont replacement. Biological indispensability of the symbiont to the host may also be relevant to symbiont replacement. For example, symbiont replacements have repeatedly occurred in weevils (Lefevre et al., 2004; Conord et al., 2008; Toju et al., 2010), and experimental studies show for several weevil lineages that symbionts are certainly beneficial, but not essential

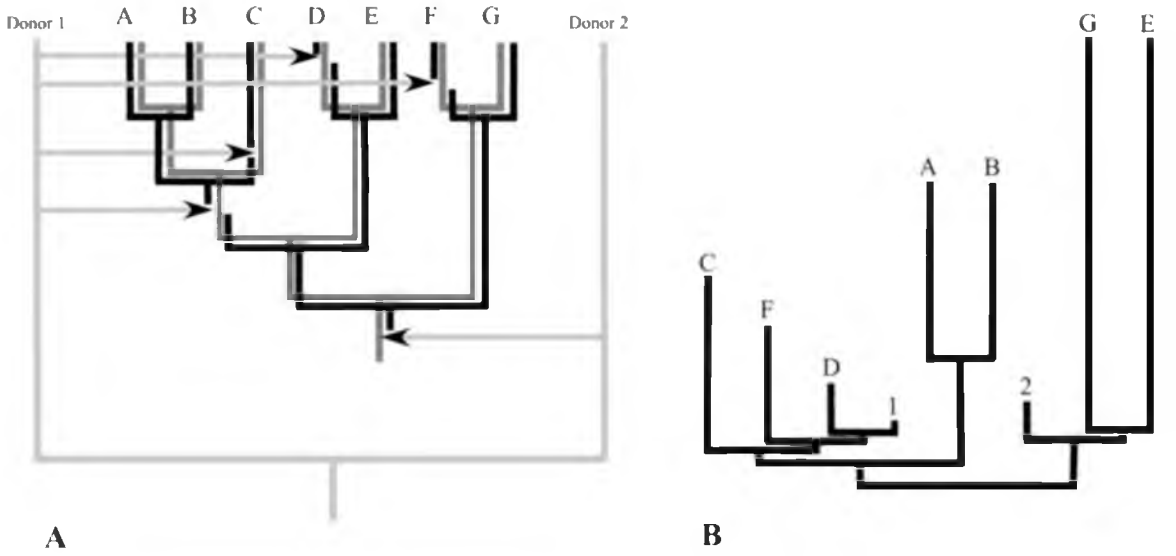


Figure C.9. Hypothetical host (grey lines) and symbiont (black lines) phylogenies generated under the symbiont replacement model. Panel A depicts a progenitor lineage (Donor 2) that colonized an ancestral host. In some cases, Donor 2 has more recently been replaced by Donor 1 (a phylogenetically distinct progenitor; replacement indicated by symbiont branch shifting from right to left of host branch). Donor 1 may also replace Donor 1 descended lineages (no branch shift indicated). The resulting symbiont phylogeny is shown in Panel B. Lineages that recently colonized hosts are more closely related to the Donor 1 progenitor than are lineages that colonized hosts in the more distant past. The model also assumes that DNA substitution rates, as indicated by branch length, increase in symbiotic lineages, relative to their free-living progenitors. As a result, branch length is proportional to the time that a given symbiont has resided in its insect host. Over time, multiple replacements results in a lack of concordance between host and symbiont phylogenies, with short internodes (reflecting the low relative rate of donor sequence evolution) connecting terminal long branches (reflecting the increased relative rate of symbiont sequence evolution), resulting in a comb- or star-like symbiont phylogeny.

for host survival. Indeed, symbiont-free weevils can grow and reproduce, but suffer retarded growth, smaller body size and/or lower flight ability (Heddi and Nardon, 2005; Kuriwada et al., 2010). It seems plausible, although speculative, that symbiont genome degeneration and indispensability jointly affect evolutionary trajectories leading to symbiont replacement. Among functionally indispensable symbionts, losses and replacements are difficult to reconcile, and these symbionts tend to co-speciate with their

hosts and evolve towards highly reduced genomes. By contrast, among beneficial but not essential symbionts, symbiont loss should be more readily tolerated, facilitating opportunities for symbiont replacements.

In conclusion, we have demonstrated unexpected diversity and evolutionary dynamics of the bacterial symbionts in feather lice of the genus *Columbicola*. To account for the peculiar evolutionary patterns observed in *Columbicola*-symbiont associations, we propose a hypothesis of repeated symbiont acquisition from a common "progenitor" bacterial lineage present somewhere in the environment. We also point out that symbiont genome degeneration and symbiont indispensability may affect evolutionary patterns of symbiont replacements. The polyphyletic bacterial symbionts of *Columbicola* spp. highlight the diversity and complexity of insect-microbe symbiotic systems, and provide insights into how such symbiotic associations have established and diversified.

Acknowledgements

The authors thank M. Reed, S. Bush, and J. Allen for various forms of assistance. We also thank T. Davis, J. Weckstein, S. Deko, N. Hillgarth, M. Robbins, P. Loi and J. Kirchman for collecting some of the lice used in this study. All procedures followed guidelines of the Institutional Animal Care and Use Committee of the University of Utah. This work was supported by National Science Foundation grant DEB 0614565 to DHC and CD, DBI 0102112 and DEB 0555024 to DLR and National Science Foundation Doctoral Dissertation Improvement Grant DEB 0608329 to WAS and DHC.

References

Baumann, P. 2005. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu. Rev. Microbiol. **59**:155-189.

Bush, S. E. and D. H. Clayton. 2006. The role of body size in host specificity: Reciprocal transfer experiments with feather lice. Evolution **60**:2158-2167.

Bush, S. E., R. D. Price and D. H. Clayton. 2009. Descriptions of eight new species of feather lice in the genus *Columbicola* (Phthiraptera: Philopteridae), with a comprehensive world checklist. J. Parasitol. **95**:286-294.

Cannone J. J., S. Subramanian, M. N. Schnare, J. R. Collett, L. M. D'Souza, Y. Du, B. Feng, N. Lin, L. V. Madabusi, K. M. Müller, N. Pande, Z. Shang, N. Yu and R. R. Gutell. 2002. The Comparative RNA Web (CRW) Site: An Online Database of Comparative Sequence and Structure Information for Ribosomal, Intron, and Other RNAs. BioMed. Central Bioinformatics **3**:2.

Clayton, D. H. and D. M. Drown. 2001. Critical evaluation of five methods for quantifying chewing lice (Insecta: Phthiraptera). J. Parasitol. **87**:1291-1300.

Clayton, D. H., R. J. Adams and S. E. Bush. 2008. Phthiraptera, the Chewing Lice. Pp. 515-526 in C. T. Atkinson, N. J. Thomas, and D. B. Hunter (eds). *Parasitic diseases of wild birds*. Wiley-Blackwell, Ames, Iowa.

Conord, C. L., A. Despres, A. Vallier, A. Balmand, C. Miquel, S. Zundel, G. Lemperiere and A. Heddi. 2008. Long-term evolutionary stability of bacterial endosymbiosis in curculionoidea: additional evidence of symbiont replacement in the dryophthoridae family. Mol. Biol. Evol. **25**:859-868.

Dale, C, and I. Maudlin. 1999. *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. Int. J. Syst. Bacteriol. **49**:267-275.

Dale, C., G. R. Plague, B. Wang, H. Ochman and N.A Moran. 2002. Type III secretion systems and the evolution of mutualistic endosymbiosis. PNAS **99**:12397-12402.

Douglas, A. E. 1989. Mycetocyte symbiosis in insects. Biol. Rev. Camb. Philos. Soc. **64**:409-434.

Edgar, R. C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics **5**:113

Farris, J. S., M. Källersjö, A. G. Kluge and C. Bult. 1994. Testing significance of incongruence. Cladistics **10**:315-319.

Farris, J. S., M. Källersjö, A. G. Kluge and C. Bult. 1995. Constructing a significance test for incongruence. Syst. Biol. **44**:570-572.

Fukatsu T., S. Aoki, U. Kurosu, and H. Ishikawa. 1994. Phylogeny of Cerataphidini

aphids revealed by their symbiotic microorganisms and basic structure of their galls: Implications for host-symbiont coevolution and evolution of sterile soldier castes. Zoological Science **11**:613-623.

Fukatsu, T., and H. Ishikawa. 1992. A novel eukaryotic extracellular symbiont in an aphid, *Astegopteryx styraci* (Homoptera, Aphididae, Hormaphidinae). J. Insect Physiol. **38**:765-773.

Fukatsu, T. and H. Ishikawa. 1996. Phylogenetic position of yeast-like symbiont of *Hamiltonaphis styraci* (Homoptera, Aphididae) based on 18S rDNA sequence. Insect Biochem. Mol. Biol. **26**:383-388.

Fukatsu, T., R. Koga, W. A. Smith, K. Tanaka, N. Nikoh, K. Sasaki-Fukatsu, K. Yoshizawa, C. Dale and D. H. Clayton. 2007. Bacterial endosymbiont of the slender pigeon louse *Columbicola columbae*, allied to endosymbionts of grain weevils and tsetse flies. App. Environ. Microbiol. **73**:6660-6669.

Gillespie, J. M., and M. J. Frenkel. 1974. The diversity of keratins. Comp. Biochem. Physiol. **47B**: 339-346.

Grunwald, S., M. Pilhofer, and W. Höll. 2010. Microbial associations in gut systems of wood- and bark-inhabiting longhorned beetles [Coleptera: Cerambycidae]. Syst. Appl. Microbiol. **33**:25-34.

Gruwell, M. E., N. B. Hardy, P. J. Gullan, and K. Dittmar. 2010. Evolutionary relationships among primary endosymbionts of the mealybug subfamily Phenacoccinae (Hemiptera: Coccoidea: Pseudococcidae). Appl. Environ. Microbiol. **76**:7521-7535.

Guindon, S. and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. **52**:696-704.

Harbison, C. W., S. E. Bush, J. R. Malenke and D. H. Clayton. 2008. Comparative transmission dynamics of competing parasite species. Ecology **89**: 3186-3194.

Harbison, C. W. and D. H. Clayton. 2011. Community interactions govern host switching with implications for host-parasite coevolutionary history. PNAS **108**:9525-9529.

Heddi, A., and P. Nardon. 2005. *Sitophilus oryzae* L: a model for intracellular symbiosis in the Dryophthoridae weevils (Coleoptera). Symbiosis **39**:1-11.

Hoy, M. A. and A. Jeyaprakash. 2005. Microbial diversity in the predatory mite *Metaseiulus occidentalis* (Acari: Phytoseiidae) and its prey, *Tetranychus urticae* (Acari: Tetranychidae). Biol. Control **32**:427-441.

Huelsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics **17**:754-755.

Huigens, M.E., R. P. de Almeida, P. A. Boons, R. F. Luck, and R. Stouthamer. 2004. Natural interspecific and intraspecific horizontal transfer of parthenogenesis-inducing *Wolbachia* in *Trichogramma* wasps. Proc. R. Soc. London Ser. B **271**:509-515.

Hypsa, V. and J. Krizek. 2007. Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (Phthiraptera, Anoplura). Microb. Ecol. **54**:242-251.

Jaenike J, M. Polak, A. Fiskin, M. Helou, and M. Minhas. 2007. Interspecific transmission of endosymbiotic *Spiroplasma* by mites. Biol. Lett. **3**:23-25.

Johnson, K. P. and D. H. Clayton. 2003. The biology, ecology, and evolution of chewing lice. pp 449 – 476 In The Chewing Lice: World Checklist and Biological Overview, R. D. Price, R. A. Hellenthal, R. L. Palma, K. P. Johnson, and D. H. Clayton. Illinois Natural History Survey Special Publication 24.

Johnson, K. P., D. L. Reed, S. Hammond, D. Kim and D. H. Clayton. 2007. Phylogenetic analysis of nuclear and mitochondrial genes supports species groups for *Columbicola* (Insecta: Phthiraptera). Mol. Phyl. Evol. **45**:506-518.

Johnson, K. P. and J. D. Weckstein. 2011. The Central American land bridge as an engine of diversification in New World doves. Journal of Biogeography **38**:1069-1076.

Kaiwa, N., T. Hosokawa, Y. Kikuchi, N. Nikoh, X. Y. Meng, N. Kimura, M. Ito, M and T. Fukatsu. 2010. Primary gut symbiont and secondary, Sodalis-allied symbiont of the Scutellerid stinkbug. Appl. Environ. Microbiol. **76**:3486-3494.

Kaiwa, N., T. Hosokawa, Y. Kikuchi, N Nikoh, X. –Y Meng, N. Kimura, M. Ito, T. Fukatsu. 2011. Bacterial symbionts of the giant jewel stinkbug *Eucoryssus grandis* (Hemiptera: Scutelleridae). Zool. Sci. **28**:169-174.

Kuriwada T., T. Hosokawa, N. Kumano, K. Shiromoto, D. Haraguchi, and T. Fukatsu. 2010. Biological role of *Nardonella* endosymbiont in its weevil host. PLoS One **5**:e13101.

Lefevre, C. H., A. Charles, B. Vallier, B. Delobel, B. Farrell and A. Heddi. 2004. Endosymbiont phylogenesis in the dryophthoridae weevils: evidence for bacterial replacement. Mol. Biol. Evol. **21**:965-973.

Malenke, J. R., K. P. Johnson and D. H. Clayton. 2009. Host specialization differentiates cryptic species of feather-feeding lice. *Evolution* 63:1427-1438.

Marshall, R. C., D. F. G. Orwin and J. M. Gillespie. 1991. Structure and biochemistry of mammalian hard keratin. Electron Microsc. Rev. **4:**47-83.

McCutcheon, J. P, and C. D. von Dohlen. 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr. Biol. **21**:1366-1372.

Moran, N. A. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. PNAS. **93**:2873-2878.

Moran, N. A. and H. Dunbar. 2006. Sexual acquisition of beneficial symbionts in aphids. PNAS. **103**:12803-12806.

Moran, N. A., H. J. McLaughlin and R. Sorek. 2009. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. Science. **323**:379-382.

Moran, N. A., J. P. McCutcheon and A. Nakabachi. 2008. Genomics and evolution of heritable bacterial symbionts. Annu. Rev. Genet. **42**:165-190.

Moran, N. A., M. A. Munson, P. Baumann and H. Ishikawa. 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. Proc. Roy. Soc. London Ser. B. **253**:167-171.

Nakabachi A, A. Yamashita, H. Toh, H. Ishikawa, H. E. Dunbar, N. A. Moran, M. Hattori M. 2007. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. Science **314**: 267.

Novakova, E., and V. Hypsa. 2007. A new *Sodalis* lineage from blood sucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of tsetse flies symbiont *Sodalis glossinidius*. FEMS Microbiol. Lett. **269**:131-135

Page, R. D. M. 1990a. Component analysis: a valiant failure? Cladistics **6**:119-136.

Page, R. D. M. 1990b. Temporal congruence and cladistic analysis of bio- geography and cospeciation. Syst. Zool. **39**: 205–226.

Page, R. D. M. 1995. TreeMap for Macintosh, ver. 1. 0b. Available at http: //taxonomy. zoology. gla. ac. uk/rod/treemap. Html

Pei, A.Y., W. E. Oberdorf, C. W. Nossa, et al.,; (16 co-authors). 2010. Diversity of 16S rRNA genes within individual prokaryotic genomes. Appl. Environ. Microbiol. **76**:3886-3897.

Pereira, S. L., K. P. Johnson, D. H. Clayton, and A. J. Baker. 2007. Mitochondrial and nuclear DNA sequences support a Cretaceous origin of Columbiformes and dispersal-driven radiation in the Paleogene. Syst. Biol. **56:**656-672.

Posada, D. 2008. jModelTest: Phylogenetic Model Averaging. Molecular Biology and Evolution **25**:1253-1256.

Posada, D. and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics **14**:817-818.

Ries, E. 1931. Die Symbiose der Lause und Federlinge. Z. Morphol. Oekol. Tiere **20**:233-367.

Russell J. A., A. Latorre, B. Sabater-Muñoz, A. Moya, and N. A. Moran. 2003. Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. Mol. Ecol. **12**:1061-1075.

Sakurai, M., R. Koga, T. Tsuchida, X. Y. Meng and T. Fukatsu. 2005. *Rickettsia* symbiont in the pea aphid *Acyrthosiphon pisum*: novel cellular tropism, effect on host fitness, and interaction with the essential symbiont *Buchnera.* Appl. Environ. Microbiol. **71**:4069-4075.

Shimodira, H. and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. **16**:1114-1116.

Swofford, D. L. 2003. PAUP**. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Thao, M. L., N. A. Moran, P. Abbot, E. B. Brennan, D. H. Burckhardt, and P. Baumann. 2000. Cospeciation of psyllids and their primary prokaryotic endosymbionts. Appl Environ. Microbiol. **66**:2898-905.

Thao, M.L., P. J. Gullan, and P. Baumann. 2002. Secondary (gamma-Proteobacteria) endosymbionts infect the primary (beta-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. Appl. Environ. Microbiol. **68**:3190-3197.

Toju H., T. Fukatsu. 2011. Diversity and infection prevalence of endosymbionts in natural populations of the chestnut weevil: relevance of local climate and host plants. Mol. Ecol. **20**:853-868.

Toju, H., T. Hosokawa, R. Koga, N. Nikoh, X. Y. Meng, N. Kimura and T. Fukatsu. 2010. "*Candidatus* Curculioniphilus buchneri," a novel clade of bacterial endocellular symbionts from weevils of the genus *Curculio*. Appl. Environ. Microbiol. **76**:275-282.

Waterhouse, D. F. 1957. Digestion in insects. Ann. Rev. Entomol. **2**:1-18.

Weinert, L.A., J. H. Werren, A. Aebi, G.N. Stone, and F. M. Jiggins. 2009. Evolution

and diversity of *Rickettsia* bacteria. BMC Biol. **7**:6.

Wernegreen, J. J. 2002. Genome evolution in bacterial endosymbionts of insects. Nat. Rev. Genet. **3**:850-861.

Werren, J. H., W. Zhang, and L. R. Guo. 1995. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. Proc. R. Soc. London Ser. B **261**:55-63.
Wuyts, J., Y. Van de Peer and R. De Wachter. 2001. Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. Nucl. Acids Res. **29**:5017-5028.

Wuyts, J., Y. Van de Peer and R. De Wachter. 2001. Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. Nucl. Acids Res. **29**:5017-5028.